8

The Santa Fe Model for Limit Order Books

Done properly, computer simulation represents a kind of "telescope for the mind", multiplying human powers of analysis and insight just as a telescope does our powers of vision. With simulations, we can discover relationships that the unaided human mind, or even the human mind aided with the best mathematical analysis, would never grasp.

(Mark Buchanan)

In this chapter, we generalise the single-queue model that we first encountered in Chapter 5 to now account for the dynamics of all queues in an LOB. As we will discuss, the resulting model is simple to formulate but quickly leads to difficult mathematics. By implementing a wide range of mathematical techniques, and also some simplifications, we illustrate how the model is able to account for several important qualitative properties of real LOBs, such as the distribution of the bid–ask spread and the shape of the volume profiles.

We also discuss how some of the model's assumptions fail to account for important empirical properties of real LOBs, such as the long-range correlations in order flow that we discuss in Chapter 10, and we note that the model leads to conditions that allow highly profitable market-making strategies, which would easily be spotted, exploited and eliminated in real markets. We therefore argue that more elaborate assumptions motivated by empirical observations must be included in the description at a later stage. Still, considering the model in the simple form that we study throughout this chapter has important benefits, including understanding the mathematical frameworks necessary to address the many interacting order flows and eliciting the degree of complexity of the corresponding dynamics.

8.1 The Challenges of Modelling LOBs

Despite the apparent simplicity of the rules that govern LOBs, building models that are both tractable and useful has proven to be an extremely difficult task. We begin this chapter by highlighting some of the challenges that make LOB modelling so difficult.

8.1.1 State-Space Complexity

One key difficulty with modelling LOBs is that their set of possible states is so large. Measured in units of the lot size υ_0, the volume at each price level in an LOB can take any integer value (positive for buys, negative for sells). Therefore, if a given LOB offers N_P different choices for price, then the state space of the volume profile is $(N_P+1)\mathbb{N}^{N_P}$, since there are N_P+1 different possible positions for the mid-price separating buys from sells. In addition, the constraint that buy and sell orders should not overlap introduces a non-linear coupling between the two types of orders, thus forbidding the use of standard linear methods (see Chapter 19 for a model that explicitly accounts for this constraint).

This huge state space is an example of the so-called *curse of dimensionality*, which affects the modelling and analysis of high-dimensional systems in many different fields. The curse of dimensionality creates important difficulties for LOB modelling: even given huge quantities of LOB data, the number of independent data points that correspond to a specified LOB state is often very small. Therefore, performing robust calibration of an LOB model from empirical data can be extremely difficult unless the dependence on the state of the LOB is specified in a parsimonious way. A key objective common to many LOB models is to find a way to simplify the evolving, high-dimensional state space, while retaining an LOB's most important features.

8.1.2 Feedback and Control Loops

The state of an LOB clearly depends on the order flow generated by the market participants trading within it. This order flow is not static but instead fluctuates considerably through time. One key ingredient that drives the order flow generated by the market participants – especially liquidity providers – is the state of the LOB itself. Therefore, the temporal evolution of an LOB is governed by complex, dynamic feedback and control loops between order flow and liquidity.

One example of such a dependency is illustrated by the flow lines in Figure 7.6, which captures the influence of the relative volume at the best quotes on their subsequent average evolution. Writing expressions that capture this feedback in a general LOB model is a difficult task.

Another key difficulty with many stylised LOB models is that they can end up in a state in which the full LOB is completely depleted. By contrast, the feedback and control loops in real markets tend to cause the LOB to replenish as the total number of limit orders becomes small. This makes the probability of observing a real LOB with no limit orders extremely small. However, this situation is not impossible, because the stabilising feedback loops sometimes break down. Such situations can lead to price jumps and flash crashes. This is a very important theme that we will discuss in the very last chapter of this book, Chapter 22.

8.2 The Santa Fe Model

We now describe the LOB model that we will consider in this chapter, and that will serve as the foundation for many of our subsequent discussions. In a nutshell, the model generalises the single-queue model that we considered in Chapter 5 by allowing orders to reside at any price on a discrete grid. We call this model the **Santa Fe model** because it was initially proposed and developed by a group of scientists then working at the Santa Fe Institute.[1]

Consider the continuous-time temporal evolution of a set of particles on a doubly infinite, one-dimensional lattice with mesh size equal to one tick ϑ. Each location on the lattice corresponds to a specified price level in the LOB. Each particle is either of type A, which corresponds to a sell order, or of type B, which corresponds to a buy order. Each particle corresponds to an order of a fixed size υ_0, which we can arbitrarily set to 1 (see Figure 8.1). Whenever two particles of opposite type occupy the same point on the pricing grid, an annihilation $A + B \to \emptyset$ occurs, to represent the matching of a buy order and a sell order in $\mathscr{L}(t)$. Particles can also evaporate, to represent the cancellation of an order by its owner. As above, $a(t)$ is the ask-price, defined by the position of the leftmost A particle, and $b(t)$ is the bid-price, defined by the position of the rightmost B particle. As always, the mid-price is $m(t) = (b(t) + a(t))/2$.

In this **zero-intelligence** model, order flows are assumed to be governed by the following stochastic processes,[2] where all orders have size $\upsilon_0 = 1$:

- At each price level $p \leq m(t)$ (respectively $p \geq m(t)$), buy (respectively sell) limit orders arrive as a Poisson process with rate λ, independently of p.
- Buy and sell market orders arrive as Poisson processes, each with rate μ.
- Each outstanding limit order is cancelled according to a Poisson process with rate ν.
- All event types are mutually independent.

[1] This group of scientists, led by J. D. Farmer, published several papers on their model throughout the early 2000s; see Section 8.9 for detailed references.

[2] In this chapter, we will only consider the symmetric case where buy and sell orders have the same rates. The model can be extended by allowing different rate parameters on the buy and sell sides of the LOB.

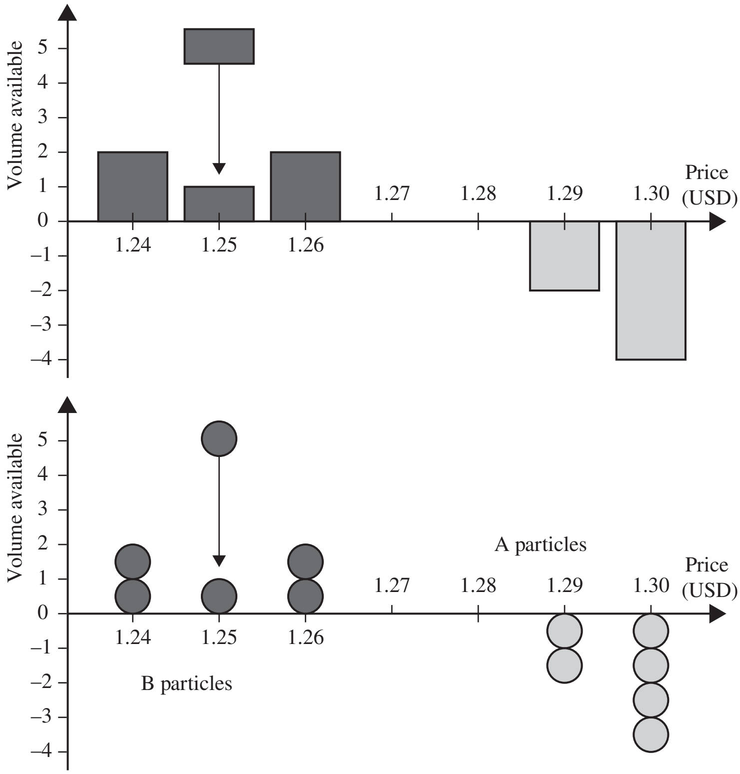

Figure 8.1. (Top) An LOB $\mathscr{L}(t)$ at some moment in time, and (bottom) its corresponding representation as a system of particles on a one-dimensional pricing lattice.

The limit order rule means that the mid-price $m(t)$ is the reference price around which the order flow organises. Whenever a buy (respectively, sell) market order x arrives, it annihilates a sell limit order at the price $a(t_x)$ (respectively, buy limit order at the price $b(t_x)$), and thereby causes a transaction. The interacting flows of market order arrivals, limit order arrivals, and limit order cancellations together fully specify the temporal evolution of $\mathscr{L}(t)$.

When restricted to the queues at the bid- and ask-prices, this model is exactly equivalent to the linear model described by Equation 7.2, with a total cancellation rate that is proportional to V^{ζ} with $\zeta = 1$. The model is also the same at all other prices, but without the possibility of a market order arrival (so $\mu = 0$ at all prices except for the best quotes).

8.3 Basic Intuitions

Before investigating the model in detail, we first appeal to intuition to discuss three of its more straightforward properties. First, each of the parameters λ, μ and

ν are rate parameters, with units of inverse time. Therefore, in order for the units to cancel out correctly, any observable quantity related to the equilibrium distribution of volumes, spreads or gaps between filled prices can only depend on ratios of λ, μ and ν. This is indeed the case for the formulae that we describe in the remainder of this section.

Second, the approximate distributions of queue sizes can be derived by considering the interactions of the different types of order flows at different prices. Because market order arrivals only influence activity at the best quotes, and because queues do not interact with one another, it follows that very deep into the LOB, the distribution of queue sizes reaches a stationary state that is independent of the distance from $m(t)$. For $V \in \mathbb{N}$ (where V is in units of the lot size υ_0), this distribution is given by Equation (5.35) with $\mu = 0$:

$$P_{\text{st.}}(V) = e^{-V^*} \frac{V^{*V}}{V!}, \qquad V^* = \frac{\lambda}{\nu}. \tag{8.1}$$

Two extreme cases are possible:

- A **sparse LOB**, corresponding to $V^* \ll 1$, where most price levels are empty while the others are only weakly populated. This case corresponds to the behaviour of the LOB for very small-tick assets.
- A **dense LOB**, corresponding to $V^* \gg 1$, where all price levels are populated with a large number of orders. This corresponds to the behaviour of the LOB for large-tick assets (at least close enough to the mid-price so that the assumption that λ is constant is reasonable).

In real LOBs, λ decreases quite steeply with increasing distance from $m(t)$ (see Figure 4.5). Therefore, even in the case of large-tick stocks, we expect a crossover between a densely populated LOB close to the best quotes and a sparse LOB far away from them.[3]

The distribution in Equation (8.1) does not hold for prices close to $m(t)$. If d denotes the distance between a given price and $m(t)$, then for smaller values of d, it becomes increasingly likely that a given price was actually the best price in the recent past. Correspondingly, limit orders are depleted not only due to cancellations but also due to market orders that have hit that queue. Heuristically, one expects that for $\lambda > \mu$, the average size of the queues at a distance d from the mid-price is given by

$$V^* \approx \frac{\lambda - \mu\phi_{\text{eff}}(d)}{\nu},$$

where $\phi_{\text{eff}}(d)$ is the fraction of time during which the corresponding price level was the best quote in the recent past (of duration ν^{-1}, beyond which all memory is lost).

[3] This does not, however, imply that there are no buyers or sellers who wish to trade at prices far away from the current mid-price. As we will argue in Chapter 18, most of the corresponding liquidity is *latent*, and only becomes revealed as $m(t)$ changes.

This formula interpolates between Equation (5.46) for $d = 0$ and Equation (8.1) for d large and says that queues tend to be smaller on average in the immediate vicinity of the mid-price, simply because the market order flow plays a greater role in removing outstanding limit orders. One therefore expects that the average depth profile is an increasing function of d, at least close to $d = 0$ (where the limit order arrival rate can be considered as a constant). This is indeed what we observed in empirical data of LOB volume profiles (see Section 4.7). We discuss a simple formula that describes the approximate shape of an LOB's volume profile in Section 8.7, which confirms the above intuition.

Next, we consider the size of the spread $s(t)$. For large-tick stocks, the bid- and ask-queues will both typically be long and $s(t)$ will spend most of the time equal to its smallest possible value of one tick, $s(t) = \vartheta$. For small-tick stocks, for which $V^* \lesssim 1$, the spread may become larger. Introducing the notation $\widehat{s}$ for the spread (measured in ticks), the probability per unit time that a new limit order (either buy or sell) arrives inside the spread can be estimated as $(\widehat{s} - 1)\lambda$. The probability per unit time that an order at one of the best quotes is removed by cancellation or by an incoming market order is given by $2(\mu + \nu)$, since the most probable volume at the bid or at the ask is $V = 1$. The *equilibrium spread* size is such that these two effects compensate,

$$s_{\text{eq.}} \approx \vartheta\left[1 + 2\frac{\mu + \nu}{\lambda}\right].$$

Although hand-waving, this argument gives a good first approximation of the average spread. Indeed, the simple result that a large rate μ of market orders opens up the spread sounds reasonable. We will return to a more precise discussion of spread sizes in Section 8.6.3.

8.4 Parameter Estimation

We now turn to the task of fitting the Sante Fe model to empirical data. At first sight, it appears that estimating most of the model's parameters should be extremely straightforward. However, one of the model's simplifying assumptions causes a considerable difficulty for model fitting. Specifically, the model assumes that the values of λ and ν do not vary with increasing distance from the best quotes. Estimating the (assumed constant) limit order arrival rate in the model requires dividing the total size of all limit order arrivals by the width of the price interval studied. However, this assumption is at odds with empirical data (see Section 4.4), because the rate of limit order arrivals tends to decrease with increasing d. Therefore, if this simplistic method is used to estimate λ from real data, then the resulting estimate of λ decreases with increasing width of the price interval studied.

One simple way to address this problem is to estimate λ only from the limit orders that arrive reasonably close to the current mid-price. In our implementations of the Santa Fe model, we choose to restrict our estimation of λ to the set $\mathcal{X}_{LO}$ of limit orders that arrive either at the best quotes or within the spread. We similarly write $\mathcal{X}_C$ to denote the set of order cancellations that occur at the best quotes. We write $\mathcal{X}_{MO}$ to denote the set of market orders (which, by definition, occur at the best quotes). We write N_{LO}, N_C and N_{MO} to denote, respectively, the number of limit order arrivals at the best quotes, cancellations at the best quotes, and market order arrivals, within a certain time window. We perform all counts independently of the corresponding order signs (recall that we assume symmetry between buy and sell activities).

We estimate the model's parameters as follows:

- To estimate the order size υ_0, we calculate the mean size of arriving limit orders

$$\upsilon_0 = \frac{1}{N_{LO}} \sum_{x \in \mathcal{X}_{LO}} \upsilon_x, \tag{8.2}$$

  where υ_x is the size of order x. Using the mean size of arriving market orders produces qualitatively similar results.
- To estimate the total (buy + sell) market order arrival rate per event, $2\tilde{\mu}$, we calculate the total size of arriving market orders (buy or sell), expressed in units of υ_0 and excluding hidden-order execution volume, then divide by the total number of events:

$$2\tilde{\mu} = \frac{1}{N_{MO} + N_{LO} + N_C} \sum_{x \in \mathcal{X}_{MO}} (\upsilon_x / \upsilon_0). \tag{8.3}$$

- To estimate the total limit order arrival rate per event, $2\tilde{\lambda}_{\text{all}}$, we simply divide the total number of limit orders by the total number of events:

$$2\tilde{\lambda}_{\text{all}} = \frac{N_{LO}}{N_{MO} + N_{LO} + N_C} \equiv \frac{1}{N_{MO} + N_{LO} + N_C} \sum_{x \in \mathcal{X}_{LO}} (\upsilon_x / \upsilon_0). \tag{8.4}$$

  The limit order arrival rate in the Santa Fe model is a rate per unit price, so to estimate $\tilde{\lambda}$, we divide $\tilde{\lambda}_{\text{all}}$ by the mean number n of available price levels inside the spread and at the best quotes, measured only at the times of limit order arrivals:

$$\tilde{\lambda} = \frac{\tilde{\lambda}_{\text{all}}}{n}, \tag{8.5}$$

  with

$$n := 2\left(1 + \left\langle \left\lfloor \frac{\widehat{s}}{2} \right\rfloor \,\middle|\, \text{event=LO} \right\rangle \right),$$

Table 8.1. *Estimated values of υ_0, $\tilde{\mu}$, $\tilde{\nu}$ and $\tilde{\lambda}$ for each of the ten stocks in our sample. We also give the average volume at the best quotes, and the spread $\langle s \rangle$ for each stock, sampled uniformly in time.*

	$\tilde{\lambda}$	$\tilde{\nu}$	$\tilde{\mu}$	υ_0 [shares]	$\bar{V}/\upsilon_0$	$\langle s \rangle$ [ticks]
SIRI	0.236	0.0041	0.013	2387	114.9	1.08
INTC	0.222	0.012	0.019	328	36.4	1.17
CSCO	0.229	0.012	0.014	545	38.1	1.14
MSFT	0.220	0.013	0.022	238	36.1	1.18
EBAY	0.208	0.022	0.029	168	23.4	1.21
FB	0.169	0.041	0.031	140	13.0	1.48
TSLA	0.023	0.109	0.062	103	3.1	21.4
AMZN	0.018	0.107	0.055	92	3.1	32.6
GOOG	0.014	0.118	0.049	84	3.0	39.2
PCLN	0.0037	0.132	0.033	68	2.8	156.7

where $\lfloor x \rfloor$ means the integer part of x. This value of n takes into account the fact that both buy and sell limit orders can fill the mid-price level when the spread (in ticks) is an even number.

- To estimate the total cancellation rate *per unit volume* and per event, $2\tilde{\nu}$, we proceed similarly and write:

$$2\tilde{\nu} = \frac{1}{N_{MO} + N_{LO} + N_C} \sum_{x \in \mathcal{X}_C} \frac{\upsilon_x}{\bar{V}}; \qquad \bar{V} := \frac{\bar{V}_a + \bar{V}_b}{2}. \tag{8.6}$$

Note that the above estimation procedures all lead to "rates per event" rather than "rates per unit time". In other words, we have not determined the continuous-time values of μ, ν, and λ, but rather the ratios $\mu/(\mu+\nu+\lambda)$, $\nu/(\mu+\nu+\lambda)$ and $\lambda/(\mu+\nu+\lambda)$. However, for the statistical properties of the LOB, these ratios are the important quantities. The only effect of considering their overall scale is to "play the movie" of order arrivals and cancellations at different speeds.

Table 8.1 lists the estimates of υ_0, $\tilde{\mu}$, $\tilde{\nu}$ and $\tilde{\lambda}$ for each stock in our sample.

8.5 Model Simulations

As we have already emphasised, deriving analytical results about the behaviour of the Santa Fe model is deceptively difficult. By contrast, simulating the model is relatively straightforward. Before we turn our attention to an analytical study of the model, we therefore perform a simulation study. Specifically, we estimate the model's input parameters from empirical data, then measure several of the model's outputs and compare them to the corresponding properties of the same data.

The Santa Fe model is extremely rich, so many output observables can be studied in this way. We consider the following four topics:

(i) the mean and distribution of the bid–ask spread;
(ii) the ratio between the mean first gap behind the best quote (i.e. the price difference between the best and second-best quotes) and the mean spread;
(iii) the volatility and signature plot of the mid-price;
(iv) the mean impact of a market order and the mean profit of market-making.

When simulating the Santa Fe model there is a trade-off between computation time and unwanted finite-size effects, induced by an artificially truncated LOB beyond some distance from the mid-price. A suitable choice is a system of size at least ten times larger than the average spread, which allows the system to equilibrate relatively quickly. We choose the initial state of the LOB such that each price level is occupied by exactly one limit order. Because small-tick LOBs exhibit gaps between occupied price levels, and because each price level in large-tick LOBs is typically occupied by multiple limit orders, this initial condition in fact corresponds to a rare out-of-equilibrium state whose evolution allows one to track the equilibration process. Once the system is in equilibrium, the standard event-time averages of the quantities listed at the start of this section can be computed from the synthetic Santa Fe time series, and then compared to empirical data.[4]

We will see in the following sections that the model does a good job of capturing some of these properties, but a less good job at capturing others, due to the many simplifying assumptions that it makes. For example, the model makes good predictions of the mean bid–ask spreads, but it predicts volatility to be too small for large-tick stocks and too large for small-tick stocks. Moreover, the model creates profitable opportunities for market-making strategies that do not exist in real markets. These weaknesses of the model provide insight into how it might be improved by including either some additional effects such as the long-range correlations of order flow (which we discuss in Chapter 10) or some simple strategic behaviours from market participants.

8.5.1 The Bid–Ask Spread

Figure 8.2 shows the empirical mean bid–ask spread versus the mean bid–ask spread generated by simulating the Santa Fe model with its parameters estimated from the same data, for each of the ten stocks in our sample. As the plot illustrates, the model fares quite well, but slightly underestimates the mean bid–ask spread for all of the stocks. This could be a simple consequence of our specific parameter estimation method (see Section 8.4), which only considers the arrivals

[4] Recall that throughout this chapter event-time refers to all events occurring at the best quotes.

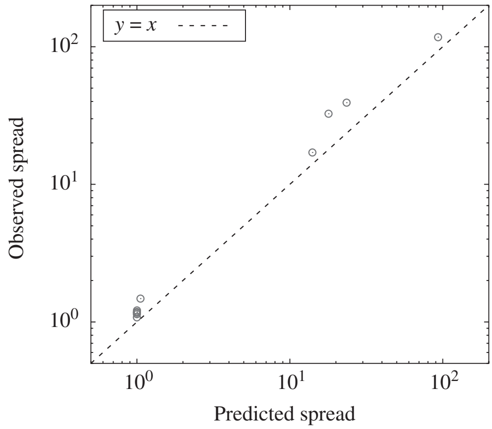

Figure 8.2. Empirical observations versus Sante Fe model predictions of the mean bid–ask spread, measured in units of the tick size, using the rate parameters in Table 8.1. The dashed line indicates the diagonal. The model does a good job at predicting the mean spread for all stocks in our sample.

and cancellations of limit orders that occur at the best quotes or inside the spread, and therefore likely causes us to overestimate the value of the limit order arrival rate.[5]

In any case, considering only the mean does not provide insight into whether the model does a good job at capturing the full distribution of the bid–ask spread. This full distribution is particularly interesting for small-tick stocks, for which the bid–ask spread can (and often does) take any of a wide range of different values. Figure 8.3 shows the distribution of the bid–ask spread for three small-tick stocks in our sample (i.e. PCLN, AMZN, TSLA), together with the corresponding results from the Santa Fe model, rescaled in such a way that the mean spread coincides in all cases. Once rescaled properly, the model does in fact reproduce quite well the bid–ask spread distribution.

8.5.2 The Gap-to-Spread Ratio

We now consider the mean gap between the second-best quote and the best quote, compared to the mean spread itself. We call this ratio the **gap-to-spread ratio**. The gap-to-spread ratio is interesting since it is to a large extent insensitive to the problem of calibrating λ correctly and of reproducing the mean bid–ask spread exactly. It is also important because this gap determines the impact of large market orders (see Equation (8.8) below).

Figure 8.4 shows the empirical gap-to-spread ratio versus the gap-to-spread ratio generated by simulating the Santa Fe model, for the same ten stocks. For large-tick

[5] Recall that the Santa Fe model also assumes that limit orders and market orders all have the same size.

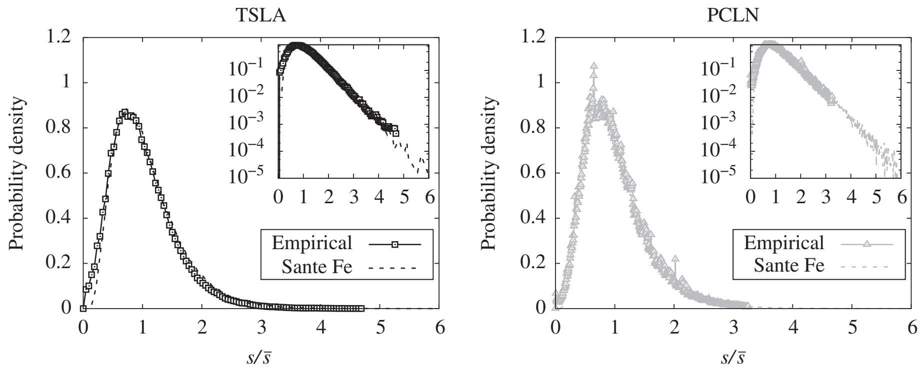

Figure 8.3. Distribution of the rescaled bid–ask spread for (left panel) TSLA and (right panel) PCLN. The markers denote the empirical values and the dashed lines denote the predictions from the Santa Fe model. The model's predictions are remarkably accurate.

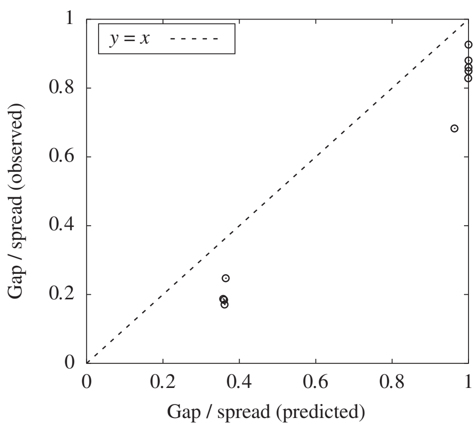

Figure 8.4. Empirical ratio of the gap between the best and second-best quote to the spread, versus the corresponding values predicted from the Santa Fe model, using the rate parameters in Table 8.1. The dashed line denotes the diagonal.

stocks, the model predicts that both the spread and the gap between the second-best and best quotes are nearly always equal to one tick, so the ratio is simply equal to 1. In reality, however, the situation is more subtle. The bid–ask spread actually widens more frequently than the Santa Fe model predicts, such that the empirical gap-to-spread ratio is in fact in the range 0.8–0.95. This discrepancy becomes even more pronounced for small-tick stocks, for which the empirical gap-to-spread ratio typically takes values around 0.2, but for which the Santa Fe model predicts a gap-to-spread ratio as high as 0.4. In other words, when the spread is large, the second-best price tends to be much closer to the best quote in reality than it is in the model.

This analysis of the gap-to-spread ratio reveals that the Santa Fe model is missing an important ingredient. When the spread is large, real liquidity providers tend to place new limit orders inside the spread, but still close to the best quotes, and thereby typically only improve the quote price by one tick at a time. This leads to gaps between the best and second-best quotes that are much smaller than those predicted by the assumption of uniform arrivals of limit orders at all prices beyond the mid-price. One could indeed modify the Santa Fe specification to account for this empirically observed phenomenon of smaller gaps for limit orders that arrive inside the spread, but doing so comes at the expense of adding extra parameters.[6]

8.5.3 Volatility

Simulating the Santa Fe model reveals that the signature plot $\sigma(\tau)/\sqrt{\tau}$ (see Section 2.1.4) exhibits a typical **mean-reversion** pattern, in particular when both the cancellation rate ν and the average size of the best queues $V^* \approx (\lambda - \mu)/\nu$ are small (see Figure 8.5).

Clearly, when ν is large, the memory of the LOB is completely erased after a short time ν^{-1}, beyond which the model has Markovian dynamics and the price is diffusive. Suppose now that ν is small and V^* is large, and that both queues are initially in equilibrium, with comparable volumes of order V^*. The first-hitting time T_1 after which one of the best queues depletes (and the mid-price changes) then grows exponentially with V^* (see Equation (5.47)). If the bid (say) manages to deplete, the empty level just created can refill with either a buy limit order or a sell limit order with equal probability. However, these two situations are not exactly symmetrical: if the new limit order is a buy, then the position of the old bid is restored, facing again the same ask-queue as before, with volume of order V^*. If however the new limit order is a sell, then the incipient ask-queue faces the previous second-best bid, which was previously shielded from the flow of market orders and is therefore typically longer than the best queues.

In other words, the initial volume imbalance I is more favourable to the old bid than to the new ask, leading to some mean reversion. The relative difference in volume imbalance is of order μ/λ, because of the shielding effect. Therefore, the effect is small unless μ is comparable to or greater than λ, in which case the LOB is sparse and V^* small (see Section 8.6). In this case, mean-reversion effects can be substantial (see Figure 8.5). In the other limit when ν is small and V^* large, the volatility of the mid-price is very small (because the depletion time T_1 grows exponentially with V^*, leading to very infrequent price changes) and the signature plot is nearly flat.

We can now turn to a comparison of the Santa Fe model's prediction about volatility with the empirical volatility of our pool of stocks. First, from the results

[6] On this point, see Mike, S., & Farmer, J. D. (2008). An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32, 200–234.

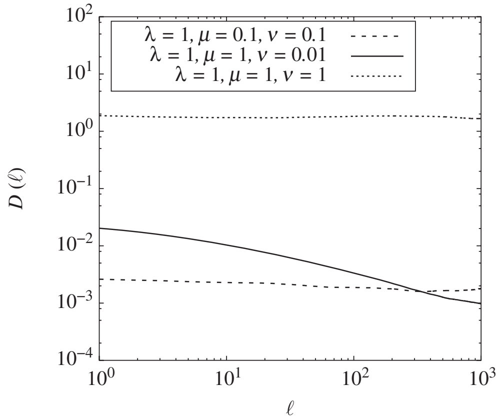

Figure 8.5. Signature plot for the Santa Fe model, with parameters (dashed curve) $\lambda = 1$ and $\nu = \mu = 0.1$, (dotted curve) $\lambda = 1$ and $\nu = \mu = 1$, and (solid curve) $\lambda = 1$, $\nu = 0.01$ and $\mu = 1$. The model exhibits mean-reversion effects only in the regime where $\nu \ll \lambda$ and $\nu \ll \mu$.

in Table 8.1, we can see that large-tick stocks are in a regime where V^* is much larger than 1. From the discussion above, we understand that the volatility predicted by the Santa Fe model is very small, and grossly underestimates the empirical volatility (see Figure 8.6). In reality, these extremely long depletion times are tempered, partly because the assumption of a constant cancellation rate per order is inadequate, as discussed in Sections 6.5 and 7.3.3.

For small-tick stocks, the model predicts values of volatility higher than those observed empirically (see Figure 8.6). The main reason for this weakness is the absence of a mechanism that accurately describes how order flows adapt when prices change. In the model, once the best quote has disappeared, the order flow immediately adapts around the new mid-price, irrespective of whether the price change was caused by a cancellation, a market order arrival or a limit order arrival inside the spread. In the language of Chapter 11, the permanent impact of all of these events is identical, whereas in reality the permanent impact of a market order arrival is much larger than that of cancellations.[7] This causes volatility in the Santa Fe model to be higher than the volatility observed empirically for small-tick stocks.

8.5.4 Impact of Market Orders and Market-Making Profitability

We now analyse in more detail the **lag-dependent impact** of market orders in the Santa Fe model. We define this lag-τ impact as

$$\mathcal{R}(\tau) := \langle \varepsilon_t \cdot (m_{t+\tau} - m_t) \rangle_t, \tag{8.7}$$

[7] On this point, see Eisler, Z., Bouchaud, J. P., & Kockelkoren, J. (2012). The price impact of order book events: Market orders, limit orders and cancellations. *Quantitative Finance*, 12(9), 1395–1419.

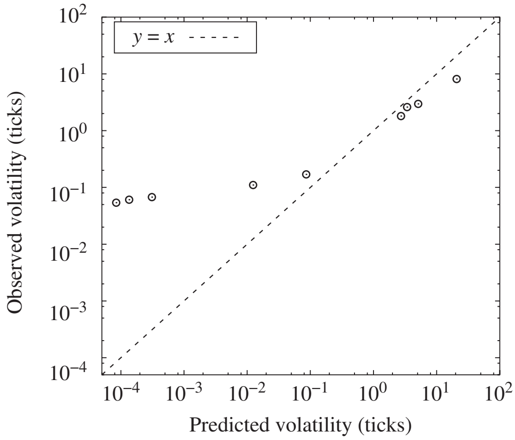

Figure 8.6. Empirically observed volatility and corresponding predictions from the Santa Fe model, for the ten stocks listed in Table 8.1. For the large-tick stocks, we estimate the volatility using the theoretical first-hitting time formula, $\sigma = \sqrt{2/T_1}$, with $T_1 = \exp(V^*)/(\nu V^*)$ and $V^* = (\lambda - \mu)/\nu$. SIRI has a predicted volatility of 10^{-12} and is not plotted here. The model overestimates the observed volatility of (rightmost points) small-tick stocks and grossly underestimates the observed volatility of (leftmost points) large-tick stocks. The dashed line denotes the diagonal.

where ε_t denotes the sign of the market order at event-time t, where t increases by one unit for each market order arrival. As we will discuss in Chapter 11, in real markets the impact function $\mathcal{R}(\tau)$ is positive and grows with τ before saturating for large τ. In the Santa Fe model, however, $\mathcal{R}(\tau)$ is constant, independent of τ, when mean-reversion effects are small (as discussed in the previous section). The lag-τ impact is then well approximated by:

$$\mathcal{R}(\tau) \approx \mathbb{P}(V_{\text{best}} = 1) \times \frac{1}{2} \langle \text{first gap} \rangle, \tag{8.8}$$

where $\mathbb{P}(V_{\text{best}} = 1)$ is the probability that the best queue is of length 1. This approximation holds because a market order of size 1 impacts the price if and only if it completely consumes the volume at the opposite-side best quote, and if it does so, it moves that quote by the size of the first gap, and thus moves the mid-price by half this amount. Since order flow in the Santa Fe model is uncorrelated and is always centred around the current mid-price, the impact of a market order is instantaneous and permanent. Therefore, the model predicts that $\mathcal{R}(\tau)$ is constant in τ for small-tick stocks. In real markets, by contrast, the signs of market orders show strong positive autocorrelation (see Chapter 10), which causes $\mathcal{R}(\tau)$ to increase with τ.

This simple observation has an important consequence: the Santa Fe model specification leads to **profitable market-making strategies**, even when the

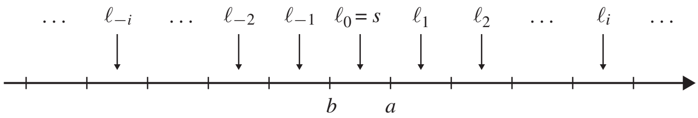

Figure 8.7. Illustration of the kinematics of the model.

signature plot is flat (i.e. when prices are diffusive). In Section 1.3.2, we argued that market-making is profitable on average if the mean bid–ask spread is larger than twice the long-term impact $\mathcal{R}_\infty$. In the Santa Fe model, the mean first gap is always smaller than the bid–ask spread. Therefore, from Equation (8.8), one necessarily has $\mathcal{R}_\infty < \langle s \rangle/2$, so market-making is easy within this framework. If we want to avoid the existence of such opportunities, which are absent in real markets, we need to find a way to extend the Santa Fe model. One possible route for doing so is incorporating some strategic behaviour into the model, such as introducing agents that specifically seek out and capitalise on any simple market-making opportunities that arise. Another route is to modify the model's assumptions regarding order flow to better reflect the empirical properties observed in real markets. For example, introducing the empirically observed autocorrelation of market order signs would increase the long-term impact $\mathcal{R}_\infty$ and thereby reduce the profitability of market-making.

8.6 Some Analytical Results

In this section, we turn our attention to deriving some analytical results from the Santa Fe model. Despite the simplicity of its modelling assumptions, analytical treatment of the full model (as specified in Section 8.2) is extremely difficult. However, as we will see throughout this section, if we adopt the assumption that the sizes of gaps in the LOB are independent, then it is possible to derive several interesting results from the model, such as the distribution of the bid–ask spread. The methods that we use to obtain these results are also interesting in their own right, and are useful in many other contexts.

Throughout this section, we use the index i to label the different intervals (i.e. gaps between successive prices at which limit orders reside) in the LOB. We use positive indices to label these gaps on the sell-side, negative indices to label these gaps on the buy-side, and the index $i = 0$ to label the bid–ask spread (see Figure 8.7).

We write ℓ_i to denote the size of gap i. Due to the finite tick size in an LOB (see Section 3.1.5), the ℓ_i must always take values

$$\ell_i \in \{0, \vartheta, 2\vartheta, \ldots\}, \qquad i = \ldots, -2, -1, 0, 1, 2, \ldots.$$

We write $P_i(\ell)$ to denote the probability distribution of ℓ_i. The value $\ell_i = 0$ corresponds to the case where the distance between the i^{th} and $(i+1)^{\text{th}}$ occupied prices is exactly one tick (i.e. there is no gap in the LOB). In this notation, we can express the bid–ask spread as $s = \ell_0 + \vartheta$ and the i^{th} gap as $\ell_i + \vartheta$. In the following, we will set $\vartheta = 1$, such that all gap sizes are measured in tick units.

As we mentioned at the start of this section, we will assume that all of the ℓ_i are *independent* random variables. Under this assumption, it is possible to write an exact recursion equation for the P_i, by the following argument. Due to the nature of order flow in the model, each queue in the LOB disappears (i.e. depletes to 0) with a certain

Poisson rate (see Section 5.4.3). Let γ_i denote this Poisson rate of disappearing for the i^{th} level. We also introduce a corresponding variable λ_i to denote the rate per unit time and per unit length that a new limit order falls in the i^{th} interval. In the Santa Fe model, λ_i is simply a constant equal to λ for all i. In this section, we consider a slight extension, where λ_i can be arbitrary, for example $\lambda_i = \lambda$ for all $i \neq 0$, but with λ_0 possibly different, to reflect that the arrival rate of limit orders inside the spread can differ considerably from the arrival rate of limit orders at other prices in the LOB.

We now consider the temporal evolution of the gap sizes. For concreteness, we discuss an interval $i \geq 1$ (i.e. on the sell-side of the LOB); similar arguments hold by symmetry for the buy-side of the LOB (the case $i = 0$ is treated separately). Several different events can contribute in changing the gap size ℓ_i between t and $t + \mathrm{d}t$:

(i) With probability $\lambda_i \mathrm{d}t \times \ell_i$, a new sell limit order arrives inside the i^{th} gap and cuts it into two intervals of size ℓ'_i and ℓ'_{i+1}, such that $\ell'_i + \ell'_{i+1} + 1 = \ell_i$ (one tick is now occupied with the new limit order). The size of the i^{th} gap is ℓ'_i. In the Santa Fe model, limit order arrivals occur at a uniform rate at all prices in the gap, so ℓ'_i is uniformly distributed between 0 and $\ell_i - 1$.

(ii) With probability $\lambda_{i-1} \mathrm{d}t \times \ell_{i-1}$, a new sell limit order arrives inside the $i-1^{th}$ gap and cuts it into two intervals of size ℓ'_{i-1} and ℓ'_i, such that $\ell'_{i-1} + \ell'_i + 1 = \ell_{i-1}$. The new size of the i^{th} gap is ℓ'_i, which is again uniformly distributed between 0 and $\ell_{i-1} - 1$.

(iii) With probability $\mathrm{d}t \times \sum_{j=0}^{i-2} \lambda_j \ell_j$, a new sell limit order arrives inside any interval to the left of $i-1$, and thereby increases the label of i by one, such that the new ℓ_i is equal to the old ℓ_{i-1}.

(iv) With probability $\gamma_{i+1} \mathrm{d}t$, the right boundary level of the i^{th} gap disappears, which causes ℓ_i to increase to $1 + \ell_{i+1}$.

(v) With probability $\sum_{j=1}^{i} \gamma_j \mathrm{d}t$, one of the queues to the left of interval i disappears, which causes the old $i+1^{th}$ interval to become the new i^{th} interval, such that the new ℓ_i is equal to the old ℓ_{i+1}.

8.6.1 The Master Equation

Assuming that all ℓ_i are independent, the above enumeration of all possible events enables us to write an evolution equation for the marginal distributions $P_i(\ell)$. Let

$$\bar{\ell}_i = \sum_{\ell=0}^{\infty} \ell \, P_i(\ell) \tag{8.9}$$

denote the mean of ℓ_i. Removing the explicit t-dependence of the P_i (to simplify the notation), one has:

$$\begin{aligned}\frac{\partial P_i(\ell)}{\partial t} &= -\lambda_i \ell_i P_i(\ell) + \lambda_i \sum_{\ell'=\ell+1}^{\infty} P_i(\ell') + \lambda_{i-1} \sum_{\ell'=\ell+1}^{\infty} P_{i-1}(\ell') \\ &\quad + \gamma_{i+1} \sum_{\ell'=0}^{\ell-1} P_i(\ell') P_{i+1}(\ell - \ell' - 1) \\ &\quad - \left(\sum_{j=1}^{i+1} \gamma_j + \sum_{j=0}^{i-1} \lambda_j \bar{\ell}_j \right) P_i(\ell) + \sum_{j=1}^{i} \gamma_j P_{i+1}(\ell) + \sum_{j=0}^{i-2} \lambda_j \bar{\ell}_j P_{i-1}(\ell)\end{aligned} \tag{8.10}$$

for all $i \geq 1$.

This equation looks extremely complicated, but is actually relatively friendly in Laplace space. Introducing

$$\widehat{P}_i(z) := \sum_{\ell=0}^{\infty} e^{-z\ell} \, P_i(\ell),$$

we arrive (after some simple manipulations) at

$$\frac{\partial \widehat{P}_i(z)}{\partial t} = \lambda_i \widehat{P}_i'(z)) + \frac{1}{1-e^{-z}}\Big(\lambda_i(1-\widehat{P}_i(z)) + \lambda_{i-1}(1-\widehat{P}_{i-1}(z))\Big) + \gamma_{i+1}e^{-z}\widehat{P}_i(z)\widehat{P}_{i+1}(z)$$
$$-\left(\sum_{j=1}^{i+1}\gamma_j + \sum_{j=0}^{i-1}\lambda_j\bar{\ell}_j\right)\widehat{P}_i(z) + \sum_{j=1}^{i}\gamma_j\widehat{P}_{i+1}(z) + \sum_{j=0}^{i-2}\lambda_j\bar{\ell}_j\widehat{P}_{i-1}(z), \tag{8.11}$$

where $\widehat{P}'(z)$ is the derivative of $\widehat{P}(z)$ with respect to z.

If we consider the case $i = 0$, we can write the equation for the distribution of the bid–ask spread. Recalling that in the Santa Fe model, order flow is symmetric on the buy and sell-sides of the LOB, and noting that both buy and sell limit and market orders can influence the bid–ask spread, we see that:

$$\frac{\partial \widehat{P}_0(z)}{\partial t} = \lambda_0\widehat{P}_0'(z) + \frac{\lambda_0}{1-e^{-z}}\Big(1-\widehat{P}_0(z)\Big) + 2\gamma_1 e^{-z}\widehat{P}_0(z)\widehat{P}_1(z) - 2\gamma_1\widehat{P}_0(z). \tag{8.12}$$

In the following we will consider the case where $\lambda_{|i|\geq 1} := \lambda$, as assumed in the original version of the model. We will determine the stationary solutions such that the left-hand side of Equations (8.11) and (8.12) are zero, both deep in the book and for the bid–ask spread.

8.6.2 The "Deep" Solution

Deep inside the LOB (i.e. when $i \gg 1$), one expects that $\gamma_i \to \gamma^*$ and that $\widehat{P}_i(z) \to Q^*(z)$. Plugging this into Equation (8.11), one finds:

$$0 = \lambda Q^{*\prime}(z) + \frac{2\lambda}{1-e^{-z}}(1-Q^*(z)) + \gamma e^{-z}Q^{*2}(z) - (\gamma^* + \lambda\bar{\ell}^*)Q^*(z), \tag{8.13}$$

where $\bar{\ell}^*$ is the mean gap length in the limit $i \to \infty$. This non-linear, ordinary differential equation admits the following solution:

$$Q^*(z) = \frac{1}{1+\bar{\ell}^*(1-e^{-z})}, \quad \text{with} \quad \bar{\ell}^* = \frac{\gamma^*}{\lambda}. \tag{8.14}$$

This corresponds to a geometric distribution

$$P^*(\ell) = \frac{1}{\ell^*}\left(1-\frac{1}{\ell^*}\right)^{\ell} \tag{8.15}$$

for the gap size ℓ, with a mean interval length equal to γ/λ.

The value of γ can be calculated from the single-queue problem that we considered in Chapter 5. At queues deep in the LOB, one can neglect the influence of the market orders and set $\mu = 0$. Equation (5.35) then leads to the stationary distribution of the volume at each occupied price:

$$P_{\text{st.}}(V) = \frac{1}{e^{\lambda/\nu}-1}\frac{\left(\frac{\lambda}{\nu}\right)^V}{V!}, \text{ for } V \geq 0.$$

The rate γ^* at which an occupied price becomes unoccupied (i.e. the total volume at the given price depletes to 0) is given by

$$\gamma^* = \nu P_{\text{st.}}(V=1),$$
$$= \frac{\lambda}{e^{\lambda/\nu}-1},$$

which finally leads to

$$\bar{\ell}^* = \frac{\gamma^*}{\lambda} = \frac{1}{e^{\lambda/\nu} - 1}.$$

In the limit $\nu \gg \lambda$ (i.e. where the cancellation rate is much higher than the deposition rate, so the LOB is sparse, as typically occurs for small-tick stocks) the mean gap size is very large, of order ν/λ. In the limit $\nu \ll \lambda$ (i.e. where the deposition rate is much higher than the cancellation rate, so the LOB is densely populated, as typically occurs for large-tick stocks), the mean gap size is very small, because the gaps are equal to zero most of the time, and are rarely equal to one tick (and even more rarely equal to two ticks, etc.). Although these two regimes lead to very different LOB states, the gap distribution is always a geometric distribution.

8.6.3 The Bid–Ask Spread

We now consider Equation (8.12) in the stationary limit, for which

$$\lambda_0 Q_0'(z) + \frac{\lambda_0}{1 - e^{-z}}\left(1 - \widehat{P}_0(z)\right) + 2\gamma_1 e^{-z}\widehat{P}_0(z)\widehat{P}_1(z) - 2\gamma_1 \widehat{P}_0(z) = 0. \tag{8.16}$$

Let us first assume that the tick size is very large, such that all gaps with $|i| \geq 1$ have a very small mean length $\bar{\ell}_i \ll 1$. As a first approximation, we can assume $\widehat{P}_1(z) \approx 1$, and thus simplify Equation (8.16) to yield

$$\widehat{P}_0(z) \approx 1 + (e^z - 1)\int_z^{\infty} \mathrm{d}z' \,\frac{e^{-d_0(z'-z+e^{z'}-e^z)} - 1}{4\sinh^2(z'/2)}; \qquad d_0 = \frac{2\gamma_1}{\lambda_0}.$$

Taking the limit $z \to 0$ of the above integral leads to a general expression for the mean bid–ask spread. Two limiting cases are interesting:

- If $d_0 \ll 1$ (which corresponds to an intense in-flow of limit orders), then the mean spread is close to one tick, and the mean gap size is given by

$$\bar{\ell}_0 \approx d_0 \ll 1.$$

  In this case, the probability that the spread is equal to exactly one tick is given by $(1 + d_0)^{-1}$.
- If $d_0 \gg 1$ (which corresponds to the case where the market order rate μ is large compared to the flow of limit orders inside the spread), then

$$\bar{\ell}_0 \approx 1.6449\ldots d_0.$$

  This is a strange (but possible) case, since we still assume that the first gap behind the best quote is nearly always equal to zero, while the spread itself is very large.

In the limit where all gaps are large (as typically occurs for small-tick stocks), the relevant range of z values is $z \ll 1$, which corresponds to the continuum limit of the model. In this case, Equation (8.16) can be solved to produce

$$\widehat{P}_0(z) = 1 - d_0 z \int_z^{\infty} \mathrm{d}z' \,\frac{1 - \widehat{P}_1(z')}{z'} e^{-d_0 \int_z^{z'} \mathrm{d}z''(1-\widehat{P}_1(z''))}. \tag{8.17}$$

Writing $\widehat{P}_0(z \to 0) \approx 1 - z\bar{\ell}_0$, one finds

$$\bar{\ell}_0 = d_0 \int_0^{\infty} \mathrm{d}z' \,\frac{1 - \widehat{P}_1(z')}{z'} e^{-d_0 \int_0^{z'} \mathrm{d}z''(1-\widehat{P}_1(z''))}.$$

In the limit $d_0 \gg 1$, this leads to an interesting, universal result that is independent of the precise shape of $\widehat{P}_1(z)$. Due to the rapid decay of the exponential term, it follows that z' must remain small, so that the result reads

$$\bar{\ell}_0 \approx d_0 \bar{\ell}_1 \int_0^\infty \mathrm{d}z' e^{-\frac{d_0}{2} z'^2 \bar{\ell}_1} = \sqrt{\frac{\pi}{2}} \times \sqrt{d_0 \bar{\ell}_1}. \tag{8.18}$$

By introducing a scaled variable $u = s\bar{\ell}_0$, one can in fact analyse the full equation for $\widehat{P}_0(z)$ in the regime $d_0 \gg 1$. In this case, Equation (8.17) leads to the result

$$\widehat{P}_0(u) = 1 - ue^{u^2/2} \int_u^\infty \mathrm{d}u' e^{-u'^2/2}.$$

The Laplace transform can be inverted and leads to the following distribution for $x = \ell_0/\bar{\ell}_0$:

$$P_0(x) = xe^{-x^2/2}.$$

The case $\mu \gg \nu \gg \lambda$ is also very interesting, because one can characterise the distribution of gap sizes for all i to find an extended power-law regime:

$$\bar{\ell}_i \approx \frac{1}{\lambda} \sqrt{\frac{\mu\nu}{|i|}}; \qquad 1 \ll |i| \ll \frac{\mu}{\nu}.$$

In particular, $\bar{\ell}_1 \sim \sqrt{\mu\nu}/\lambda$. Therefore, using $\bar{\ell}_1$, Equation (8.18) and the approximation $d_0 \approx \mu/\lambda$, we finally arrive at the scaling relation

$$\bar{\ell}_0 \sim \frac{\mu^{3/4}\nu^{1/4}}{\lambda}$$

for the mean bid–ask spread in this regime. Note that $\bar{\ell}_i \sim |i|^{-1/2}$ corresponds to a linear volume profile, since

$$L = \sum_{i=1}^{Q} \bar{\ell}_i \propto \sqrt{Q}, \tag{8.19}$$

which means that the integrated volume Q over a region of extension L away from the mid-price grows like L^2. The density of orders at distance L is therefore proportional to L.

8.7 The Continuum, Diffusive-Price Limit

Another route to understanding the shape of the LOB is to assume that the tick size is infinitesimal ($\vartheta \to 0$) so that the mid-price takes continuous values. However, working with the Santa Fe model under this assumption is difficult, because the motion of the mid-price is itself dictated by the interaction between the order flow and the instantaneous shape of the LOB. One possible way to make progress is to artificially decouple the motion of the mid-price from the state of the LOB, and to instead impose that it simply follows a random walk. As we noted in Section 8.5.3, this assumption is only valid on time scales larger than ν^{-1}, because substantial mean-reversion is present on shorter time scales.

Following this approach provides the following path to an approximate quantitative theory of the mean volume $V_{\mathrm{st.}}(d)$ at distance d from the mid-price. Buy (respectively, sell) orders at distance d from the current mid-price at time t

are those that were placed at a time $t' < t$ and have survived until time t. Between times t' and t, these orders:

(i) have not been cancelled; and
(ii) have not been crossed by the ask (respectively, bid) at any intermediate time t'', in the sense that the ask (respectively, bid) price has never been above (respectively, below) the price of the orders since the time that they were placed.

For the remainder of this section, we assume that the arrival rate of limit orders at a given price p depends on the distance d between p and the mid-price $m(t)$. Let $\lambda(d)$ denote the arrival rate of limit orders at a distance d from $m(t)$.

Consider a sell limit order in the LOB at some given time t. This order appeared in the LOB at some time $t' \leq t$, when its distance from the mid-price was $d' = d + m(t) - m(t')$. Therefore, its deposition occurred with some rate $\lambda(d')$. Using conditions (i) and (ii) above, the mean volume profile in the limit $t \to \infty$ can be written as:

$$V_{\mathrm{st.}}(d) = \lim_{t\to\infty} \int_{-\infty}^{t} \mathrm{d}t' \int_{-d}^{\infty} \mathrm{d}u \lambda (d+u) \mathbb{P}[u|t \to t'] \, \mathrm{e}^{-\nu(t-t')},$$

where $\mathbb{P}[u|t \to t']$ is the conditional probability that the time-evolution of the price produces a given value $u = m(t) - m(t')$ of the mid-price difference, given the condition that the path satisfies[8] $d + m(t) - m(t'') \geq 0$ at all intermediate times $t'' \in [t', t]$.

Evaluating $\mathbb{P}$ requires knowledge of the statistics of the price process, which we assume to be purely diffusive, with some diffusion constant D. In this case, $\mathbb{P}$ can be calculated using the standard **method of images**,[9] to yield:

$$\mathbb{P}[u|t \to t'] = \frac{1}{\sqrt{2\pi D\tau}} \left[\exp\left(-\frac{u^2}{2D\tau}\right) - \exp\left(-\frac{(2d+u)^2}{2D\tau}\right) \right],$$

where $\tau = t - t'$. After a simple computation, one finds (up to a multiplicative constant, which only affects the overall normalisation):

$$V_{\mathrm{st.}}(d) = \mathrm{e}^{-\kappa d} \int_0^d \mathrm{d}u \lambda(u) \sinh(\kappa u) + \sinh(\kappa d) \int_d^{\infty} \mathrm{d}u \lambda(u) \mathrm{e}^{-\kappa u}, \qquad (8.20)$$

where $\kappa^{-1} = \sqrt{D/2\nu}$ measures the typical range of price changes during the lifetime of an order ν^{-1}.

[8] We neglect here the size of the spread. The condition should in fact read $d + b(t) - b(t'') = d + m(t) - m(t'') - (s(t) - s(t''))/2 \geq 0$, for all $t'' \in [t', t]$.
[9] See, e.g., Redner, S. (2001). *A guide to first-passage processes*. Cambridge University Press.

Equation (8.20) depends on the statistics of the incoming limit order flow, which is modelled by the deposition rate $\lambda(d)$. Let us consider a simple case

$$\lambda(d) = e^{-\alpha d}$$

in which $\lambda(d)$ decays exponentially in d. In this case, it is possible to evaluate the integrals explicitly, to yield

$$V_{\text{st.}}(d) = V_0\left[e^{-\alpha d} - e^{-\kappa d}\right],$$

where V_0 is some volume scale. In the original specification of the Santa Fe model, $\lambda(d)$ is constant in d. In this case, $\alpha = 0$, so $V_{\text{st.}}(d) = V_0\left[1 - e^{-\kappa d}\right]$, which implies that the mean volume increases monotonically with increasing d. This is at odds with empirical observations of real LOBs, for which the mean volume profile first increases but then decreases with increasing d.

For general values of α, it is clear from the expression of $V_{\text{st.}}(d)$ that the mean volume:

- decays to 0 (linearly) when $d \to 0$. We will return to this important point in Chapter 19;
- reaches a maximum at some distance d^*; and
- decays back to 0 (exponentially) for large d.

Although only approximate, this simple framework provides insight into why the mean volume profile for small-tick stocks exhibits a universal *hump shape* (see Section 4.7 and Figure 4.8). In particular, the liquidity tends to be small close to the current mid-price. This will be a recurring theme in the following (see Chapter 18).

8.8 Conclusion: Weaknesses of the Santa Fe Model

The Santa Fe model makes many extreme assumptions that are clearly unrealistic. Some are probably innocuous, at least for a first account of the dynamics of the LOB, whereas others are more problematic, even at a qualitative level. To conclude our discussion of the model, we list some of these unrealistic assumptions and propose some possible cures.

(i) The model assumes that buy (respectively, sell) limit orders arrive at the same rate at all prices $p \leq m(t)$ (respectively, $p \geq m(t)$), whereas in real markets the rates of limit order arrivals vary strongly with relative price (see Figure 4.5). As we have seen in Section 8.7, allowing the arrival rate to vary as a function of price can be incorporated in the model relatively easily, and improves the model's ability to reproduce the behaviour observed in real LOBs, like the average volume profile.

(ii) The model assumes that all sources of temporal randomness are governed by Poisson processes with time-independent rates, whereas real order arrivals and cancellations cluster strongly in time and follow intra-day patterns (see Chapter 2). The problem of event clustering could be rectified by replacing the Poisson processes by self-exciting processes, such as Hawkes processes (see Chapter 9), and the problem of intra-day seasonalities can be eliminated by working in event-time.

(iii) The model assumes that all order flows are independent of the current state of the LOB and of previous order flow. However, as we have already seen for the dynamics of the best queues in Chapter 7, order-flow rates clearly depend both on the same-side and opposite-side best queues, and there is no reason to expect that order-flow rates will not also depend on other characteristics of the LOB. Two important dependencies can be accommodated relatively straightforwardly within an extended Santa Fe model: (a) the volume of incoming market orders tends to grow with the available volume and (b) the probability that a market order hits the ask rather than the bid increases with the volume imbalance I introduced in Section 7.2.

(iv) Perhaps even more importantly, the Santa Fe model assumes that all order flows are independent of each other. This modelling assumption is deeply flawed. As we will repeatedly emphasise in the coming chapters, there are several strong, persistent correlations between order flows in an LOB. For example, the signs of market orders are long-range autocorrelated.[10] Market order arrivals also tend to be followed by limit order arrivals that refill the depleted queue, and vice-versa, in a tit-for-tat dance between liquidity provision and consumption. This is important to produce a price that behaves approximately as a martingale.

Despite these simplifications, the model incorporates all of the concrete rules of trading via an LOB. Specifically, orders arrive and depart from the LOB in continuous time (due to particle arrivals, evaporations and annihilations), at discrete price levels (due to the discreteness of the pricing grid) and in discrete quantities (due to the discreteness of each particle's size). Whenever a particle arrives at a price already occupied by another particle of opposite type, the particles annihilate (which corresponds to the arrival of a market order). Hence, although the Santa Fe model does not seek to incorporate all empirical properties of order flow observed in real markets, it captures the essential mechanics of trading via an LOB. Therefore, the model can be regarded as a simple null model of price formation, without considering any strategic behaviour of the market participants. Interestingly, the model allows one to reproduce many empirical

[10] For models that include these long-range correlations, see: Mastromatteo, I., Tóth B., & Bouchaud, J-P. (2014). Agent-based models for latent liquidity and concave price impact. *Physical Review E*, 89(4), 042805.

quantities, such as the average or distribution of the bid–ask spread, with a remarkable accuracy, vindicating the idea that for some observables the influence of intelligence may be secondary to the influence of the simple rules governing trade. More sophisticated ingredients, such as modelling optimised strategies of market-makers/high-frequency traders or the fragmentation of large orders that lead to correlated order flows, can be included at a later stage.

Take-Home Messages

(i) Modelling the full LOB is difficult because the state space is large and investors' actions are extremely complex.

(ii) The so-called zero-intelligence approach simplifies the problem by assuming that investors' actions follow stochastic processes. Though oversimplified, this approach allows some empirical stylised facts to be recovered. This shows that some regularities of markets may simply be a consequence of their stochastic nature.

(iii) The Santa Fe model regards the LOB as a collection of queues that evolve according to a constant limit order arrival rate, a constant cancellation rate per existing limit order, and a constant market order arrival rate at the best queues.

(iv) By estimating the empirical rates for each such events, the Santa Fe model can be used to make predictions of several important statistical properties of real markets.

(v) The model makes reasonably good predictions about the mean bid–ask spread, and about the full distribution of the bid–ask spread.

(vi) The model makes reasonably good predictions about the increasing mean volume profile near to the best quotes, but does not capture the eventual decrease in mean volume deeper into the LOB.

(vii) The model slightly overestimates volatility for small-tick stocks, but grossly underestimates volatility for large-tick stocks.

(viii) In the model, the mean impact of an order is smaller than the half spread, creating profitable market-making opportunities. This is because the zero-intelligence approach misses a crucial ingredient of financial markets: feedback of prices on strategic behaviour.

8.9 Further Reading

Stylised Facts and Modelling

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.
Weber, P., & Rosenow, B. (2005). Order book approach to price impact. *Quantitative Finance*, 5(4), 357–364.
Mike, S., & Farmer, J. D. (2008). An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32(1), 200–234.
Bouchaud, J. P., Farmer, J. D., & Lillo, F. (2009). How markets slowly digest changes in supply and demand. In Hens, T. & Schenk-Hoppe, K. R. (Eds.), *Handbook of financial markets: Dynamics and evolution*. North-Holland, Elsevier.
Eisler, Z., Bouchaud, J. P., & Kockelkoren, J. (2012). The price impact of order book events: Market orders, limit orders and cancellations. *Quantitative Finance*, 12(9), 1395–1419.
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.

The Santa Fe Model

Daniels, M. G., Farmer, J. D., Gillemot, L., Iori, G., & Smith, E. (2003). Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. *Physical Review Letters*, 90(10), 108102.
Smith, E., Farmer, J. D., Gillemot, L. S., & Krishnamurthy, S. (2003). Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6), 481–514.
Farmer, J. D., Patelli, P., & Zovko, I. I. (2005). The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6), 2254–2259.

Agent-Based Models of LOBs

Bak, P., Paczuski, M., & Shubik, M. (1997). Price variations in a stock market with many agents. *Physica A: Statistical Mechanics and its Applications*, 246(3–4), 430–453.
Chiarella, C., & Iori, G. (2002). A simulation analysis of the microstructure of double auction markets. *Quantitative Finance*, 2(5), 346–353.
Challet, D., & Stinchcombe, R. (2003). Non-constant rates and over-diffusive prices in a simple model of limit order markets. *Quantitative Finance*, 3(3), 155–162.
Preis, T., Golke, S., Paul, W., & Schneider, J. J. (2006). Multi-agent-based order book model of financial markets. *EPL (Europhysics Letters)*, 75(3), 510.
Chiarella, C., Iori, G., & Perello, J. (2009). The impact of heterogeneous trading rules on the limit order book and order flows. *Journal of Economic Dynamics and Control*, 33(3), 525–537.
Tóth, B., Lemperiere, Y., Deremble, C., De Lataillade, J., Kockelkoren, J., & Bouchaud, J. P. (2011). Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2), 021006.
Mastromatteo, I., Tóth, B., & Bouchaud, J. P. (2014). Agent-based models for latent liquidity and concave price impact. *Physical Review E*, 89(4), 042805.
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.
Muni Toke, I. (2015). Stationary distribution of the volume at the best quote in a Poisson order book model. arXiv preprint arXiv:1502.03871.

Mathematical Models for the Dynamics of LOBs

Luckock, H. (2003). A steady-state model of the continuous double auction. *Quantitative Finance*, 3(5), 385–404.

Foucault, T., Kadan, O., & Kandel, E. (2005). Limit order book as a market for liquidity. *Review of Financial Studies*, 18(4), 1171–1217.

Rosu, I. (2009). A dynamical model of the limit order book. *The Review of Financial Studies*, 22(11), 4601–4641.

Cont, R., & De Larrard, A. (2012). Order book dynamics in liquid markets: Limit theorems and diffusion approximations. https://ssrn.com/abstract=1757861.

Abergel, F., & Jedidi, A. (2013). A mathematical approach to order book modeling. *International Journal of Theoretical and Applied Finance*, 16(05), 1350025.

Slanina, F. (2013). *Essentials of econophysics modelling*. Oxford University Press.

Horst, U., & Kreher, D. (2015). A weak law of large numbers for a limit order book model with fully state dependent order dynamics. arXiv preprint arXiv:1502.04359.

Gao, X., Dai, J. G., Dieker, T., & Deng, S. (2016). Hydrodynamic limit of order book dynamics. *Probability in the Engineering and Informational Sciences*, 1–30.

Huang, W., & Rosenbaum, M. (2015). Ergodicity and diffusivity of Markovian order book models: A general framework. arXiv preprint arXiv:1505.04936.

Muni Toke, I., & Yoshida, N. (2016). Modelling intensities of order flows in a limit order book. *Quantitative Finance*, 17, 1–19.