7

Joint-Queue Dynamics for Large-Tick Stocks

What is the logic of all this? Why is the smoke of this pipe going to the right rather than to the left? Mad, completely mad, he who calculates his bets, who puts reason on his side!

(Alfred de Musset)

In Chapter 6, we considered how to model either the best bid-queue or the best ask-queue in an LOB for a large-tick stock. Although this approach enabled us to derive many interesting and useful results about single-queue dynamics, it is clear that the dynamics of the bid-queue and the dynamics of the ask-queue mutually influence each other. In this chapter, we therefore extend our framework to consider the joint dynamics of the best bid- and ask-queues together.

We begin the chapter by discussing the bid- and ask-queues' "race to the bottom", which is the core mechanism that drives price changes in LOBs for large-tick stocks. We present some empirical results to illustrate how such races play out in reality, then turn our attention to modelling such races. We first consider a race in which the bid- and ask-queues are assumed to evolve independently, and derive several results from the corresponding model. We next turn our attention to a more general case, in which we allow the dynamics of the bid- and ask-queues to be correlated. We end the chapter by introducing an even more versatile framework, by extending the Fokker–Planck framework from the previous two chapters to also encompass the coupled dynamics of the two queues. We illustrate that this approach provides a consistent description of the joint dynamics observed in real LOBs for large-tick stocks.

This chapter is not essential to the main story of the book.

7.1 The Race to the Bottom

Recall from Chapter 3 that in an LOB for a large-tick stock, the bid–ask spread is very often equal to its minimum possible size, $s(t) = \vartheta$. This has an important consequence for LOB dynamics, because it removes the possibility that a new limit order will arrive inside the spread. Therefore, whenever $s(t) = \vartheta$, the only way for either the bid-price or the ask-price to change is for one of the order queues at the best quotes to deplete to $V = 0$.

If the bid-queue depletes before the ask-queue, then the bid-price moves down (but might bounce back up immediately – see below). If the ask-queue depletes before the bid-queue, then the ask-price moves up (but might bounce back down immediately). Therefore, price changes in this situation can be regarded as the consequence of a simple race between the bid- and ask-queues, in which the winner (i.e. the queue that depletes first) dictates the direction of the next price change.

By considering some initial volumes V_b^0 and V_a^0, then considering the subsequent market order arrivals, limit order arrivals, and cancellations that occur at $b(t)$ and $a(t)$, we can derive models for the two-dimensional trajectory $(V_b(t), V_a(t))$ (see Figure 7.1), in which $V_b(t)$ and $V_a(t)$ may be regarded as racing to zero, until at some time t_1 one queue empties and the spread momentarily widens. After t_1, a limit order will very likely rapidly fill the (now widened) spread. What happens in that case will be discussed in depth in Section 7.5.

Many interesting questions about the dynamics of the bid- and ask-queues spring to mind. At some randomly chosen time t:

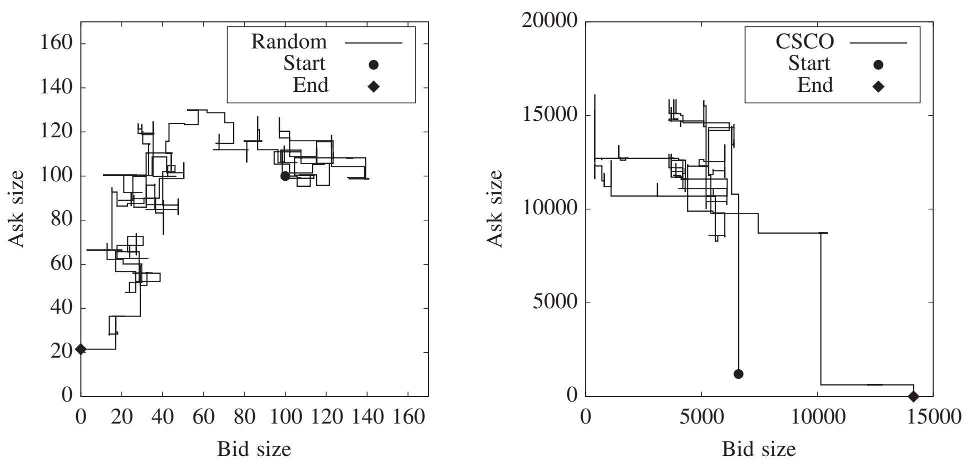

Figure 7.1. Typical trajectories in the plane (V_b, V_a), starting (circle) at (V_b^0, V_a^0) and ending (diamond) by the depletion of one of the two queues. The left panel shows simulation data for a pair of uncoupled random walks; the right panel shows the real data for CSCO.

- What is the joint probability $P(V_b, V_a)$ that the ask volume is $V_a(t) = V_a$ and the bid volume is $V_b(t) = V_b$?
- Having observed the two volumes $(V_b(t), V_a(t))$, what is the probability $p_\pm$ that the next mid-price move is upwards?
- Having observed the two volumes $(V_b(t), V_a(t))$, what is the rate $\Phi(\tau)$ for the next mid-price move (up or down) to take place at time $t+\tau$?
- Having observed the two volumes $(V_b(t), V_a(t))$, what is the probability that an order with a given priority at the bid or at the ask will be executed before the mid-price changes?

Throughout the remainder of this chapter, we address the first two topics and some other questions about joint-queue dynamics for large-tick stocks. The last two questions, as well as many others, can be studied using similar tools.

7.2 Empirical Results

We first present some empirical results to illustrate joint-queue dynamics for large-tick stocks in real markets. Given that the state space we consider is the two-dimensional space (V_b, V_a), one possible approach could be to estimate empirically the fraction of times that a race ended by an upward price movement (rather than a downward price movement) for a given initial choice of (V_b, V_a). However, this approach suffers the considerable drawback of needing to estimate and interpret a two-dimensional surface, for which many choices of initial conditions have relatively few data points because they occur relatively infrequently in real markets. Furthermore, the **scale invariance** property of the queue dynamics, noted in the previous chapter and approximately valid when queues are large, suggests to investigate this problem not by considering the two-dimensional input quantity (V_b, V_a), but instead by considering the one-dimensional **queue imbalance** ratio

$$I := \frac{V_b}{V_b + V_a}. \tag{7.1}$$

The quantity I measures the (normalised) imbalance between V_b and V_a, and thereby provides a quantitative assessment of the relative strengths of buying and selling pressure in an LOB. Observe that I is invariant upon rescaling both volumes by the same factor α (i.e. $V_a \to \alpha V_a$ and $V_b \to \alpha V_b$), and takes values in the open interval $(0,1)$. A value $I \approx 0$ corresponds to a situation where the ask-queue is much larger than the bid-queue, which suggests that there is a net positive selling pressure in the LOB that will most likely push the price downwards. A value $I \approx 1$ corresponds to a situation where the bid-queue is much larger than the ask-queue, which suggests that there is a net positive buying pressure in the LOB that will most likely push the price upwards. A value $I \approx \frac{1}{2}$ corresponds to a situation in

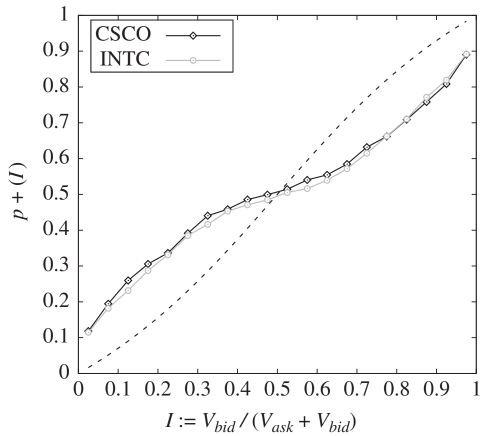

Figure 7.2. Probability that the ask-queue depletes before the bid-queue, as a function of the queue imbalance I for INTC and CSCO. We also show (dashed curves) the result of a simple model where the volume of the bid- and ask-queues undergo independent random walks, given by Equation (7.6).

which the bid- and ask-queues have approximately equal lengths, which suggests that the buying and selling pressures in the LOB are approximately balanced.

We now turn our attention to an empirical analysis of how the volume imbalance I impacts the outcome of a given race. To do so, we consider the value of I at some times chosen uniformly at random, then fast-forward to see which of the bid- or ask-queues depleted to zero first. We can then estimate the probability $p_+(I)$ as the fraction of races for which the subsequent price movement was upwards. We plot these empirical results in Figure 7.2. The graph reveals a strong, monotonic correlation between the queue imbalance and the direction of the next price move.[1]

Throughout the remainder of this chapter, we will introduce several different models for the joint-queue dynamics of large-tick stocks. For each such model, we will estimate the relevant parameters from data, then plot the model's predicted output for the probability that a race with a given input imbalance will result in an upward price movement. We will then compare each model's outputs with our empirical results in Figure 7.2 to assess how well (or otherwise) each approach is able to reproduce realistic market dynamics.

7.3 Independent Queues

We first consider a race in which the bid- and ask-queues evolve independently, according to the dynamics that we first introduced in Chapter 5 (in which limit order arrivals, market order arrivals, and cancellations are each assumed to occur

[1] This has been called the "worst kept secret of high-frequency traders", see, e.g., http://market-microstructure.institutlouisbachelier.org/uploads/91_3%20STOIKOV%20Microstructure_talk.pdf

according to independent Poisson processes). In Section 5.3, we studied a model in which cancellations occur as a Poisson process with fixed rate for the whole queue together, and in Section 5.4, we studied a model in which each individual limit order is cancelled with some rate (such that at time t, the total cancellation rate for the whole bid-queue is proportional to $V_b(t)$ and the total cancellation rate for the whole ask-queue is proportional to $V_a(t)$). For describing the dynamics of either the bid-queue or ask-queue in isolation, these approaches can be written:

$$\begin{cases} V \to \quad V+1 & \text{with rate} \quad \lambda \mathrm{d}t \\ V \to \quad V-1 & \text{with rate} \quad (\mu + \nu V^{\zeta})\mathrm{d}t, \end{cases} \tag{7.2}$$

where $\zeta = 0$ for the first model and $\zeta = 1$ for the second model.

To describe the joint dynamics of the bid- and ask-queues together, we replace the parameters λ, μ, and ν with the parameters λ_b, λ_a, μ_b, μ_a, ν_b and ν_a, where the subscript denotes the relevant queue. To begin with, let us suppose that the initial bid- and ask-queue volumes are (V_b, V_a), that their subsequent evolution is governed by (7.2), with no coupling between the two queues, and that the rate parameters obey

$$\lambda_b = \lambda_a,$$

$$\mu_b = \mu_a,$$

$$\nu_b = \nu_a.$$

When queues are independent, there are several different approaches for obtaining the probability p_+ that the ask-queue empties first. One such approach is to consider an integral that counts all the possible outcomes where the first-hitting time of the bid-queue is larger than that of the ask-queue,

$$p_+(V_b, V_a) = \int_0^\infty \mathrm{d}\tau \Phi(\tau, V_a) \int_\tau^\infty \mathrm{d}\tau' \Phi(\tau', V_b), \tag{7.3}$$

where $\Phi(\tau, V)$ is the distribution of first-hitting times considered in Section 5.3.2.

An equivalent approach is to follow a similar reasoning to the one that we employed in Equation (5.16):

$$\begin{aligned} p_+(V_b, V_a) = \frac{W_+(V_b)}{Z(V_b, V_a)} p_+(V_b+1, V_a) + \frac{W_+(V_a)}{Z(V_b, V_a)} p_+(V_b, V_a+1) \\ + \frac{W_-(V_b)}{Z(V_b, V_a)} p_+(V_b-1, V_a) + \frac{W_-(V_a)}{Z(V_b, V_a)} p_+(V_b, V_a-1), \end{aligned} \tag{7.4}$$

where $Z(V_b, V_a) = W_+(V_b) + W_-(V_b) + W_+(V_a) + W_-(V_a)$, and with the boundary conditions $p_+(V_b > 0, V_a = 0) = 1$ and $p_+(V_b = 0, V_a > 0) = 0$. We will exploit this general framework in simple cases in the next sections.

7.3.1 Races between Diffusive Queues

We first consider the diffusive case, where cancellations occur as a Poisson process with fixed rate for the whole queue together. This corresponds to the case $\zeta = 0$ in Equation (7.2) and the case $W_+(V) = \lambda$ and $W_-(V) = \mu + \nu$ in Equation (7.4), for both the bid- and the ask-queues.

As we discussed in Chapter 5, the non-critical case (where $\lambda < \mu + \nu$) is not very interesting, because it cannot produce long queues. Recall from Section 5.3.2 that the mean first-hitting time of a single queue is a Gaussian random variable with mean $V/(\mu + \nu - \lambda)$ and variance proportional to V. If one neglects fluctuations (of order $\sqrt{V}$), it is clear that the shortest queue empties first. There is significant uncertainty only if the width of the first-hitting time distribution becomes larger than the difference $|V_a - V_b|/(\mu + \nu - \lambda)$. For large queues, this can only happen when $V_a \sim V_b$, with $|V_a - V_b|$ of the order of $\sqrt{V_a}$ ($\sim \sqrt{V_b}$). In other words, the probability that the ask-queue empties first is close to 1 when $V_a \ll V_b$ and close to 0 when $V_a \gg V_b$, with a sharp, step-like transition of width $\sqrt{V_{a,b}}$ around $V_a = V_b$.

The more interesting case is the critical case (where $\lambda = \mu + \nu$). We study this case in the continuum limit, (see Section 5.3.4).[2] In this case, $W_+(V) = W_-(V) = \lambda$ and Equation (7.4) becomes **Laplace's equation** in the first quadrant ($V_b > 0, V_a > 0$):

$$\lambda\left(\partial^2_{V_bV_b} + \partial^2_{V_aV_a}\right) p_+(V_b, V_a) = 0; \qquad p_+(V_b, 0) = 1; \qquad p_+(0, V_a) = 0. \tag{7.5}$$

One approach to solving this equation is to introduce polar coordinates (R, θ), with

$$R = \sqrt{V_b^2 + V_a^2},$$
$$\theta = \arctan(V_b/V_a).$$

In polar coordinates, Laplace's equation becomes

$$\left(\partial^2_{RR} + R^{-2}\partial^2_{\theta\theta}\right) p_+(R, \theta) = 0.$$

One can check that the only acceptable solution to this equation is of the form $p_+ = A\theta$ (i.e. independent of R). The constant A is fixed by the boundary conditions, so that the solution finally reads:

$$p_+(I) := p_+(V_b, V_a) = \frac{2}{\pi} \arctan\left(\frac{I}{1 - I}\right), \tag{7.6}$$

with

$$\frac{I}{1 - I} = \frac{V_b}{V_a}. \tag{7.7}$$

Therefore, we find that p_+ is scale invariant, i.e. it depends only on the ratio V_b/V_a (or equivalently on the volume imbalance I), but not on V_a and V_b separately.

[2] For a detailed discussion of the discrete case, see Cont, R. & De Larrard, A. (2013). Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1), 1–25.

As a sanity check, any solution must obey $p_+(I) = p_-(1-I) = 1 - p_+(1-I)$, by the symmetry of the system. We can verify that Equation (7.6) indeed satisfies these conditions by direct substitution. Figure 7.2 compares Equation (7.6) with empirical results. Clearly, the independent, diffusive queue model is unable to predict the observed concavity of $p_+(I)$.

This framework can be extended to the case where $\lambda_b \neq \lambda_a$, where both queues are still critical but one queue is more active than the other. The Laplace equation then becomes:

$$\begin{aligned}\left(\lambda_b \partial^2_{V_b V_b} + \lambda_a \partial^2_{V_a V_a}\right) p_+(V_b, V_a) &= 0;\\ p_+(V_b, 0) &= 1;\\ p_+(0, V_a) &= 0.\end{aligned}$$

By setting

$$\begin{aligned} V_a &= \sqrt{\lambda_a} u_a,\\ V_b &= \sqrt{\lambda_b} u_b,\end{aligned}$$

we arrive back at the same system as in Equation (7.5), leading to:

$$p_+(V_b, V_a) = \frac{2}{\pi} \arctan\left(\sqrt{\frac{\lambda_a}{\lambda_b}} \frac{V_b}{V_a}\right).$$

This result provides a simple interpretation of the dynamics in the critical case ($\lambda = \mu + \nu$): for queues of the same size, the fastest-changing queue is more likely to empty first. For example, if $V_a = V_b$ but $\lambda_a > \lambda_b$, then $p_+ > 1/2$ (see Figure 7.3).

7.3.2 Diffusive Queues with Correlated Noise

As we saw in Section 7.3.1, p_+ is an increasing function of I, even when the order flow is completely random – meaning that short queues tend to disappear first. This empirical behaviour is well known by traders, which causes it to induce a feedback loop between traders' observations and actions. For example, when the bid-queue increases in size, traders might think that the price is more likely to go up, and therefore some sellers with limit orders in the ask-queue might cancel, while some buyers might rush to buy quickly (before the anticipated price move). Both of these actions lead to a decreased volume of the ask-queue.

Through this feedback mechanism, the dynamics of the bid- and ask-queue become negatively correlated (see Section 7.4.2). How might we extend our models of joint-queue dynamics to reflect this empirical behaviour? The simplest approach is to consider a diffusive framework with correlated noise acting on the

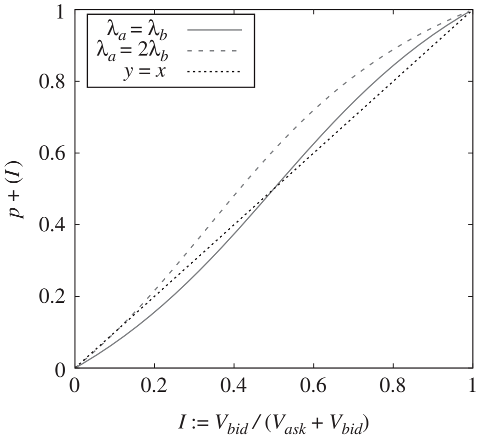

Figure 7.3. Probability that the ask-queue depletes before the bid-queue, as a function of the queue imbalance I, for diffusive queues with (solid line) $\lambda_a = \lambda_b$ and (dashed line) $\lambda_a = 2\lambda_b$. For similar volumes, the fastest changing queue has a higher probability of depleting first. The dotted line represents the diagonal.

two queues. More formally, this amounts to saying that the random increments can be written as

$$\delta V_b = \xi\sqrt{1+\rho} + \widehat{\xi}\sqrt{1-\rho},$$
$$\delta V_a = \xi\sqrt{1+\rho} - \widehat{\xi}\sqrt{1-\rho},$$

where ξ and $\widehat{\xi}$ are independent noise terms with zero mean and arbitrary but equal variances, and ρ is the correlation coefficient between the increments at the bid- and ask-queues.

From this decomposition, one sees that the sum

$$S = V_a + V_b$$

and the difference

$$\Delta = V_b - V_a$$

evolve independently, but with noises of different variances (respectively, $1+\rho$ for S and $1-\rho$ for Δ). In the (S,Δ) plane (with $S \geq 0$ and $-S \leq \Delta \leq S$), the Laplace equation from Section 7.3.1 becomes

$$\left((1+\rho)\partial^2_{SS} + (1-\rho)\partial^2_{\Delta\Delta}\right)p_+(S,\Delta) = 0, \tag{7.8}$$

with boundary conditions

$$p_+(S,S) = 1,$$
$$p_+(S,-S) = 0.$$

Defining

$$S' = \frac{S}{\sqrt{1+\rho}},$$
$$\Delta' = \frac{\Delta}{\sqrt{1-\rho}},$$

we recover exactly the independent queue problem of Section 7.3.1. In these new coordinates, the boundaries $\Delta = \pm S$ now have slopes $\pm\sqrt{(1+\rho)/(1-\rho)}$, so the wedge has an opening half-angle

$$\psi = \arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\right).$$

In terms of the new coordinates, the initial conditions read

$$S' = \frac{V_a + V_b}{\sqrt{1+\rho}},$$
$$\Delta' = \frac{V_b - V_a}{\sqrt{1-\rho}}.$$

Therefore, the angle corresponding to the initial condition (V_b, V_a) is given by

$$\theta_0 = \arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\,\frac{V_b - V_a}{V_b + V_a}\right). \tag{7.9}$$

The solution $p_+ = A\theta$ easily generalises for a wedge of arbitrary half-angle ψ, as:

$$p_+ = \frac{1}{2} + \frac{\theta_0}{2\psi}. \tag{7.10}$$

In the limit of zero correlations, $\rho = 0$, so $\psi = \pi/4$ and one indeed recovers (after simple trigonometric manipulations)

$$p_+ = \frac{2}{\pi}\arctan(V_b/V_a),$$

which is the same result as we found in Equation (7.6).

In the limit of perfectly anti-correlated queues, $\rho = -1$, so the wedge becomes extremely sharp $(\theta_0, \psi \to 0^+)$ and the problem becomes the famous one-dimensional **gambler's ruin**, for which the behaviour is linear in the imbalance:

$$p_+ = I.$$

Figure 7.4 shows p_+ as a function of I, for the three values $\psi = \pi/4$, $\psi = 3\pi/8$ and $\psi = 0^+$, which correspond to $\rho = 0$, $\rho = 0.71$ and $\rho = -1$, respectively. For all

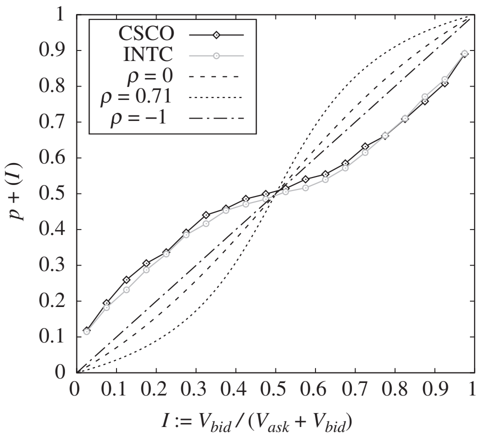

Figure 7.4. Probability that the ask-queue depletes before the bid-queue, as a function of the queue imbalance I. The solid curves denote empirical estimates for INTC and CSCO; the other curves denote theoretical predictions for (dotted curve) $\rho = 0$, (dashed curve) $\rho = 0.71$ and (dash-dotted curve) $\rho = -1$.

values of correlations, the dependence of p_+ on I has an inverted concavity when compared with empirical data. We now turn to a model that correctly captures this concavity.

7.3.3 Races between Long, Self-Stabilising Queues

We now consider the Q-CIR case (see Section 5.4.4), with $W_+(V) = \lambda$ and $W_-(V) = \mu + \nu V$, in the limit of long equilibrium sizes $V^* = (\lambda - \mu)/\nu \gg 1$. In the notation of Section 7.2, this corresponds to the case $\zeta = 1$.[3] Using the fact that the distribution of first-hitting times is Poissonian (see Section 5.4.3), Equation (7.3) becomes very simple:

$$p_+(V_b, V_a) \approx \frac{\mathbb{E}_b[T_1]}{\mathbb{E}_a[T_1] + \mathbb{E}_b[T_1]},$$

where $\mathbb{E}_b[T_1]$ and $\mathbb{E}_a[T_1]$ denote the mean first-hitting times of the bid- and ask-queues, respectively.

This result illustrates that the queue with the smaller mean first-hitting time has a larger probability of emptying first. However, because the mean first-hitting time is nearly independent of the initial queue lengths when V^* is large, we discover that in the present framework, the probability that a given queue empties first only depends very weakly on the queue imbalance I, except when at least one of the queues become small.

Figure 7.5 shows how p_+ depends on I, with an initial condition chosen with the stationary measure. Interestingly, the plateau region for $I \approx 1/2$ is less pronounced

[3] In fact, the results that we present here actually hold for any $\zeta > 0$.

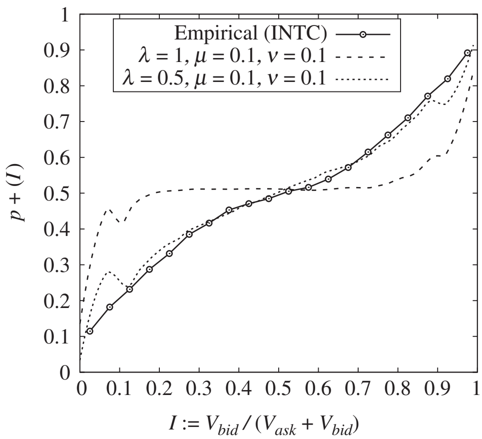

Figure 7.5. Probability that the ask-queue depletes before the bid-queue, as a function of the queue imbalance I within the stationary Q-CIR model with (dashed line) $\lambda = 1$ and $\mu = \nu = 0.1$ (corresponding to $V^* = 10$), and (dotted line) $\lambda = 0.5$ and $\mu = \nu = 0.1$ (corresponding to $V^* = 5$). We also show the same empirical data as in Figure 7.2 for INTC; its concavity is well captured by the Q-CIR model with a moderate value of V^*.

when V^* is not too large, and the curves start resembling quite closely those observed empirically. In particular, the Q-CIR model (i.e. with $\zeta = 1$) provides qualitative insights into the concavity of $p_+(I)$ that are not correctly predicted by the diffusive model (i.e. with $\zeta = 0$).

We thus recover the same conclusion as in Section 6.5: the queue dynamics of large-tick stocks is in a regime where drift effects are comparable to diffusion effects – in other words, that the effective potential defined in Sections 5.4.5 and 6.2.1 is shallow. Reality is probably best described by a value of ζ somewhere between 0 and 1 (see also Figure 6.1).

7.4 The Coupled Dynamics of the Best Queues

In the previous section, we postulated some simple dynamics for the volumes of the queues. We now consider a more flexible approach to modelling the joint-queue dynamics, which more clearly elicits the nature of the coupling between the two queues, and can be calibrated on empirical data.

7.4.1 A Two-Dimensional Fokker–Planck Equation

Generalising the arguments from Section 6.2, one finds (neglecting all price-changing events) that the bivariate distribution of the bid- and ask-queue sizes obeys a **two-dimensional Fokker–Planck equation**:

$$\begin{aligned}\frac{\partial P(V_b,V_a,t)}{\partial t} = &- \partial_{V_b}\left[F_b(V_b,V_a)P(V_b,V_a,t)\right] \\ &- \partial_{V_a}\left[F_a(V_b,V_a)P(V_b,V_a,t)\right] \\ &+ \partial^2_{V_bV_b}\left[D_{bb}(V_b,V_a)P(V_b,V_a,t)\right] \\ &+ \partial^2_{V_aV_a}\left[D_{aa}(V_b,V_a)P(V_b,V_a,t)\right] \\ &+ 2\partial^2_{V_bV_a}\left[D_{ba}(V_b,V_a)P(V_b,V_a,t)\right], \end{aligned} \tag{7.11}$$

where $\vec{F} := (F_b, F_a)$ is a two-dimensional vector field called the **flow field** and $\mathbf{D} = (D_{aa}, D_{ab}; D_{ba}, D_{bb})$ is a space-dependent, two-dimensional **diffusion matrix**. Similarly to Equation (6.8), these objects are defined as:

$$\begin{aligned} F_i(V_b,V_a) &:= \frac{1}{\tau_c}\sum_{\delta V_i} \delta V_i\, \varrho(\delta V_i|V_b,V_a), \qquad i \in \{b,a\}; \\ D_{ij}(V_b,V_a) &:= \frac{1}{2\tau_c}\sum_{\delta V_i,\delta V_j} \delta V_i \delta V_j\, \varrho(\delta V_i,\delta V_j|V_b,V_a), \qquad i,j \in \{b,a\}, \end{aligned} \tag{7.12}$$

where ϱ is the distribution of elementary volume changes, conditioned to the current size of the queues.

Note that D_{ba} and D_{ab} are the only terms that require explicit knowledge of the joint distribution of simultaneous size changes in both queues. (By symmetry, these terms are equal). These are the only terms that are sensitive to correlation between the order flow at the best bid- and ask-queues. However, even if these terms were absent, the dynamics of the two queues are coupled by the explicit dependence of the drift field $\vec{F}$ and the diagonal elements of $\mathbf{D}$ on the sizes of both the bid- and ask-queues. Such a drift-induced coupling is completely absent from the models in Sections 7.3 and 7.3.2.

7.4.2 Empirical Calibration

It is interesting to consider how the flow field $\vec{F} = (F_b, F_a)$ behaves empirically. We calculate the statistics of the flow using $\delta n = 10$ events (i.e. we measure the queue lengths once every 10 events, where each event corresponds to whenever one of the two queues changes size). Figure 7.6 shows $\vec{F}$ for INTC and CSCO, as a function of rescaled volumes $u_b = V_b/\bar{V}_b$, $u_a = V_a/\bar{V}_a$. As the figure illustrates, the flow field is very similar in both cases. As we expect from **buy/sell symmetry**, the flow field obeys

$$F_b(V_b,V_a) \cong F_a(V_a,V_b).$$

These plots reveal that when both queues are small, their volumes tend to grow on average, and when both queues are large, their volumes tend to decrease on

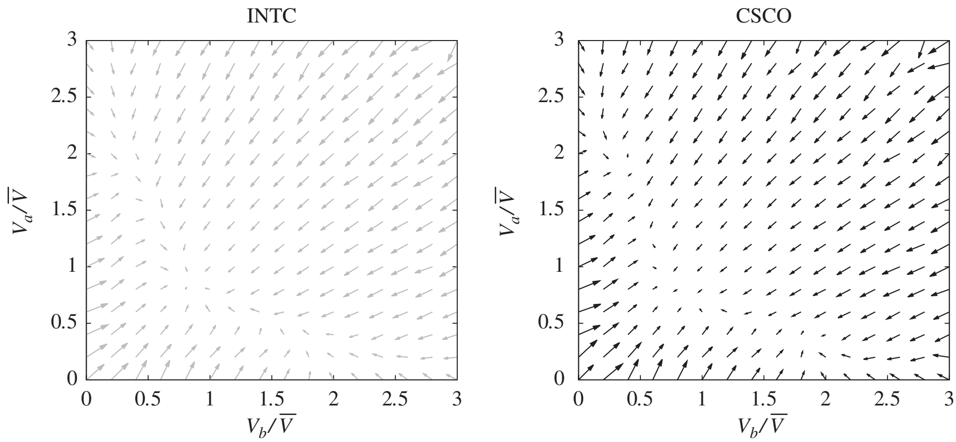

Figure 7.6. Flow field $\vec{F}$ of volume changes at the best bid and ask for (left panel) INTC and (right panel) CSCO. To improve readability, the displayed vector norms correspond to the square-root of the original vector lengths, rescaled by a factor 10.

average. The flow vanishes (i.e. $|\vec{F}| = 0$) at only one point $u_a^* = u_b^* \cong 1$. The flow lines are quite regular and appear to be nearly **curl-free**, which implies that $\vec{F}$ can be approximately written as a pure gradient $\vec{F} = -\vec{\nabla} U$, where U has a funnel-like shape pointing towards (u_b^*, u_a^*).

Of course, a significant dispersion of the dynamics around the average flow pattern occurs through the diffusion matrix. By symmetry, one expects

$$D_{bb}(V_b, V_a) = D_{aa}(V_a, V_b),$$

which is indeed well-obeyed by the data.

Empirically, the variance of volume changes is found to depend mostly on the same queue and very little on the opposite queue, so

$$D_{bb}(V_b, V_a) \cong D(V_b),$$
$$D_{aa}(V_b, V_a) \cong D(V_a),$$

where the function $D(V)$ is as in Section 6.4. The cross-diffusion terms $D_{ba} = D_{ab}$ are found to be negative and of amplitude $\cong 20\%$ of the diagonal components. This negative correlation was expected from the arguments given in Section 7.3.2, but is rather weak.

7.4.3 Stationary State

In Section 6.2.1, we saw that it is always possible to write the stationary state of the one-dimensional Fokker–Planck equation as a Boltzmann–Gibbs measure, in terms of an effective potential $\mathcal{W}(V)$. Unfortunately, this is not the case in higher dimensions (including the present two-dimensional case), except in special cases. For example, when

$$D_{aa}(V_a, V_b) = D_{bb}(V_a, V_b) = D(V_a, V_b),$$
$$D_{ab} = D_{ba} = 0,$$

and provided the flow can be written as a gradient,

$$F_{a,b}(V_a, V_b) := -D(V_a, V_b)\partial_{a,b}\mathcal{W}(V_a, V_b),$$

then the stationary state (again neglecting any reinjection term) can be written as a Boltzmann–Gibbs measure

$$P_{\text{st.}}(V_a, V_b) = \frac{A}{D(V_a, V_b)} e^{-\mathcal{W}(V_a, V_b)}. \tag{7.13}$$

This result cannot be exactly applied to data, because D_{aa} and D_{bb} are not equal to each other for all V_a, V_b. Still, because the flow field is nearly curl-free (as noted above), Equation (7.13) is an acceptable approximation in the regime where V_a and V_b are not too large, such that D is approximately constant. Such an approximation immediately leads to the conclusion that $P_{\text{st.}}(V_a, V_b)$ is peaked in the vicinity of the point where the flow field $\vec{F}$ vanishes. This agrees well with empirical observations.

7.5 What Happens After a Race Ends?

To understand the full price dynamics of large-tick stocks, one has to discuss what happens at the end of the race between the two best quotes. Imagine that the bid-queue has emptied first. Two things can happen at the now vacant price level (see Section 6.1):

- With probability ϕ_0, a buy limit order immediately refills the old bid position. In this case, the mid-price reverts to its previous position (after having briefly moved down by half a tick).
- With probability $1 - \phi_0$, a sell limit order immediately refills the old bid position. In this case the mid-price has moved down by one tick; the new ask is the old bid and the new bid is the old second-best bid.

In the second case, the price moves down, but the incipient ask-queue now faces the old second-best bid-queue. The new ask-queue is therefore typically much smaller than the bid-queue. The volume imbalance I introduced in Section 7.2 would then suggest that the new ask is more likely to disappear first. This would lead to strong mean-reversion effects where the mid-price moves back up with a large probability. In the purely stochastic models considered in the previous chapters and in the next chapter, this is indeed what happens.

Therefore at least one extra ingredient must be present to counteract this mean-reversion mechanism. As the flow lines of Figure 7.6 indicate, the new ask-queue will in fact grow quickly at first, until the system equilibrates around $I \approx 1/2$. This is the spirit of the assumption made in Cont and De Larrard

(2013), where the volume of both the new ask-queue and the new bid-queue are independently chosen from the stationary distribution $P_{\text{st.}}(V)$. Such an assumption allows one to erase all memory of the past queue configuration, and automatically generates a diffusive mid-price.

7.6 Conclusion

This ends a series of three chapters on the dynamics of long queues. Here, we have mostly focused on the joint dynamics of the bid- and ask-queues, for which the Fokker–Planck formalism offers an intuitive and flexible modelling framework. As emphasised in the last chapter, the Fokker–Planck formalism can accommodate any empirical dependence of the market order/limit order/cancellations rates simultaneously on both bid- and ask-queue volumes.

Of particular interest is the outcome of the "race to the bottom" of the bid and ask, which determines the probability of the direction of the next price change. This probability is empirically found to be correlated with volume imbalance I, with a non-trivial dependence on I. One of the most interesting conclusions of the chapter is that purely diffusive models fail to describe the quantitative shape of this function, even in the presence of noise correlations. In particular, the concavity of the empirical relation is not reproduced by the theoretical formulas.

This discrepancy can be mitigated by the introduction of a drift component in the queue dynamics, whereby small queues tend to grow on average and large queues tend to shrink on average. The dynamics of real queues are well described by a (two-dimensional) Fokker–Planck equation where the drift component is moderate compared to diffusion, preventing depletion times from becoming exceedingly large. Many dynamical properties of queues can be inferred from the interesting structure of the flow field (see Figure 7.6) and the volume dependence of the diffusion tensor (Equation (7.12)), which are found to be mostly independent of the chosen large-tick stock, provided volumes are rescaled by their average values.

Take-Home Messages

(i) Because the spread is almost always locked to one tick, the price dynamics of large-tick stocks boils down to describing the dynamics of the best bid- and ask-queues. Price changes occur when a queue empties ("wins the race").

(ii) When a queue empties, either the volume bounces back and the race continues without a change in price, or an order of opposite type fills the hole and a new race begins around the new mid-price, with the previous second-best queue on the opposite side becoming the new best queue.

(iii) In an LOB, the imbalance between the lengths of the bid- and ask-queues is a strong predictor of the direction of the next price change. The probability that the next price change is in the direction of the shorter queue is a convex function of this queue imbalance.

(iv) By assuming that queues are diffusive, one can compute the theoretical probability of a price change in a given direction as a function of the queue imbalance. However, this approach leads to predicting that this probability is a concave function of imbalance (or a linear function if the bid and ask queue lengths are perfectly anti-correlated).

(v) In the case of long, self-stabilising and independent queues, the probability of a price change in a given direction no longer depends on the queue imbalance. Instead, the price moves up or down with some fixed Poisson rate.

(vi) A more flexible modelling approach is to consider the joint dynamics of the best queues in a Fokker–Planck framework that takes into account the parameter dependence on each queue length.

7.7 Further Reading

General

Feller, W. (1971). *An introduction to probability theory and its applications* (Vol. 2). John Wiley and Sons.

Van Kampen, N. G. (1983). *Stochastic processes in physics and chemistry*. North-Holland.

Risken, H. (1984). *The Fokker–Planck equation*. Springer, Berlin-Heidelberg.

Gardiner, C. W. (1985). *Stochastic methods*. Springer, Berlin-Heidelberg.

Hänggi, P., Talkner, P., & Borkovec, M. (1990). Reaction-rate theory: Fifty years after Kramers. *Reviews of Modern Physics*, 62(2), 251.

Redner, S. (2001). *A guide to first-passage processes*. Cambridge University Press.

Races and Volume Imbalances

Avellaneda, M., Reed, J., & Stoikov, S. (2011). Forecasting prices from Level-I quotes in the presence of hidden liquidity. *Algorithmic Finance*, 1(1), 35–43.

Cartea, A., Donnelly, R. F., & Jaimungal, S. (2015). Enhanced trading strategies with order book signals. https://ssrn.com/abstract=2668277.

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.

Gould, M. D., & Bonart, J. (2016). Queue imbalance as a one-tick-ahead price predictor in a limit order book. *Market Microstructure and Liquidity*, 2(02), 1650006.

Lachapelle, A., Lasry, J. M., Lehalle, C. A., & Lions, P. L. (2016). Efficiency of the price formation process in presence of high frequency participants: A mean field game analysis. *Mathematics and Financial Economics*, 10(3), 223–262.

Fokker–Planck Equation for LOBs

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

Garèche, A., Disdier, G., Kockelkoren, J., & Bouchaud, J. P. (2013). Fokker-Planck description for the queue dynamics of large-tick stocks. *Physical Review E*, 88(3), 032809.

Yang, T. W., & Zhu, L. (2016). A reduced-form model for level-1 limit order books. *Market Microstructure and Liquidity*, 2(02), 1650008.