

Trading strategies via book imbalance

The imbalance between bid and ask orders in a limit order book tends to predict trade arrivals and price movements. Alex Lipton, Umberto Pesavento and Michael Sotiropoulos calculate probabilities of price movements given the level of book imbalance, and find it can be useful for brokers' short-term optimal trading strategies

The prevalence of electronic trading has radically changed the market structure in several asset classes, most notably in equities and futures. All market participants in a public electronic venue contribute to price formation by adding and removing liquidity in a limit order book, which ranks the buy and sell orders.

By the laws of supply and demand, when there is an order book imbalance – a large disparity between the number of buys and sells – the mid-price should move accordingly. For instance, if bids exceed asks, it should go up. Similarly, the waiting time until an order is filled will go up if the imbalance means there are many ahead of a given buy or sell order. This makes it natural to model them directly as functions of the order book imbalance.

Econometric approaches to understanding this look at the relationship between order flow and prices within electronic trading venues (Bouchaud, Mezard & Potters 2002; Engle 2000; Hasbrouck 1991; Smith *et al* 2003), while reducing the price impact caused by large orders is another focus (Almgren *et al* 2005; Bouchaud, Farmer & Lillo 2008). More recently, point processes (Bacry & Muzy 2013; Stoikov, Cont & Talreja 2010), queuing theory (Cont & Larrard 2012) and agents' utility functions (Avellaneda & Stoikov 2008) have been applied to model electronic markets.

An agency broker's task is to buy or sell a given number of shares within a set time horizon and at the best possible price. When executing on a public exchange, they can either add liquidity at the so-called near side – best bid for buy orders, best ask for sells – or remove liquidity from the far side – best ask for buys, best bid for sells. Both strategies have their trade-offs. When posting passively, paying the spread is avoided but timing options are given up. When removing liquidity, the full spread is paid but the timing of the trade can be freely chosen.

This makes the relationship between price dynamics and intensity of trade arrival an important input to trading strategy. In this article, we first compute empirically average price moves and waiting times conditional on the state of the order book. Then the probabilities for favourable price moves – down for posting at the bid, up for posting at the ask – and unfavourable price moves, as well as trade occurrences, can be computed using a stochastic model for diffusion in three dimensions. Bid and ask queues, and the timing of the near-side trade arrivals, are modelled and supplied with suitable boundary conditions.

The solution of the model takes the form of a two-dimensional Fourier–Laplace expansion with fast convergence away from the boundary, the coefficients for which can be obtained by transforming boundary conditions into a matrix equation. This is a semi-analytical solution, as the corresponding matrix elements can be found analytically when the relevant boundary is approximately linear and can be found numerically otherwise. The results are calibrated to recent historical market data.

Empirical effect of book imbalance

A common intuition among traders is that the order sizes displayed at the top of the book reflect the general intention of the market. When the number of shares at the bid exceeds that at the ask, participants expect the next price movement to be upwards, and vice versa. This is why, for example, brokers executing an order on behalf of a client might be concerned about posting too much of it at the near side and displaying their intentions.

To model this intuition we define the bid–ask imbalance as:

$$I = \frac{q^b - q^a}{q^b + q^a} \quad (1)$$

where q^b and q^a are the bid and ask quantities posted at the top of the book, and we seek to model probabilities of price movements as functions of it. Positive (respectively negative) imbalance indicates an order book that is heavier on the bid (respectively ask) side.

Figure 1 shows the effect of the book imbalance on the average mid-price change, and on the waiting time until the next price change. The data shown is based on averaging the mid-price moves of Vodafone's stock price for all trading days in the first quarter of 2012. The data is bucketed according to the initial imbalance in the book and averaged over the day. Error bars are calculated by treating the daily averages of price variations as independent data points, but they are so small that they are hardly visible at the scale of the plot.

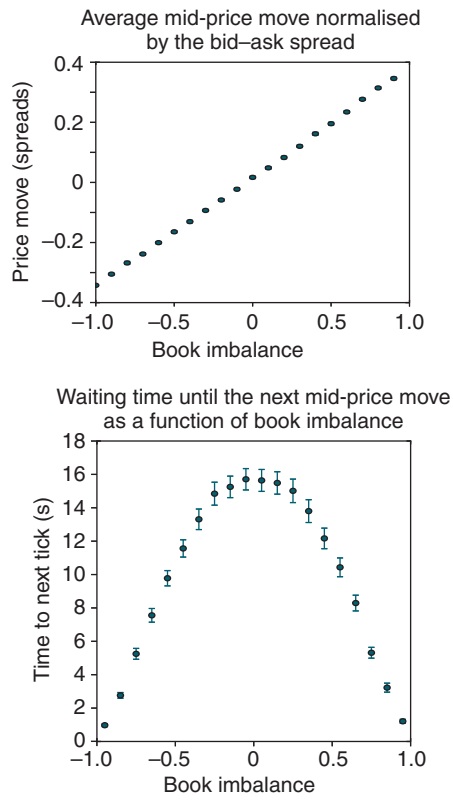
The average mid-price change is calculated by considering the stopping time defined by the next change in either the best bid or the best ask price, connecting them to the up-tick and down-tick probabilities conditional on book imbalance, as in Cont & Larrard (2012). This can be used to compute the average size of the price jump.

As expected, a high book imbalance indicates the general intention of the market and is on average a good predictor of mid-price movements. The price change until the next tick is well approximated by a linear function of the imbalance and is typically well below the bid–ask spread, even for highly imbalanced order books, therefore not offering an opportunity for a straightforward statistical arbitrage.

The average waiting time until the next mid-price change is also affected by the book imbalance, with high imbalance indicating a price move coming in a relatively short time. This relationship is expected in a market where the typical broker posts part of their orders at the near side and both price variations and queue levels are mostly driven by the pressure of the order flow.

Another stopping time with economic significance is the first arrival of a trade on a specified side. This is important to a broker posting part of their order at the near side, and having to wait for market orders originating from the far side to be matched with his. Trades from the same side may affect the share price, but will not contribute to the broker's fill rate. Therefore, for an order posted at the bid (respectively ask) side, the relevant stopping time is the time of first arrival of a sell (respectively buy) trade.

1 Effect of market imbalance on price move and waiting time



Note: both parts are based on the daily changes in the share price of Vodafone over the first quarter of 2012. The dependence of price changes on order imbalance, not just the current price, shows they are not martingales at this time scale. The average price move can be up to a third of the spread for a highly imbalanced book

A. Price moves and quote updates

t_0	p_0^b	p_0^a	q_0^b	q_0^a
t_1	p_1^b	p_1^a	q_1^b	q_1^a
t_2	p_2^b	p_2^a	q_2^b	q_2^a
\tilde{t}_0	\tilde{p}_0	\tilde{q}_0	\tilde{s}_0	
t_3	p_3^b	p_3^a	q_3^b	q_3^a
t_4	p_4^b	p_4^a	q_4^b	q_4^a
\tilde{t}_1	\tilde{p}_1	\tilde{q}_1	\tilde{s}_1	
t_5	p_5^b	p_5^a	q_5^b	q_5^a
t_6	p_6^b	p_6^a	q_6^b	q_6^a
\tilde{t}_2	\tilde{p}_2	\tilde{q}_2	\tilde{s}_2	
...

In table A we show a typical segment of the trades and quotes times series. It consists of quote vectors $(t_i, p_i^b, p_i^a, q_i^b, q_i^a)$, indicating the time of the quote update, the new best bid and ask prices and sizes, and trade vectors $(\tilde{t}_j, \tilde{p}_j, \tilde{q}_j, \tilde{s}_j)$ indicating the time, price, quantity and side of the trade. Although the side of the trade is not usually published by equities exchanges, it can be inferred from its execution price compared with the prevailing market quotes.

Average price variations and waiting times until the next buy or sell trade are calculated as follows: (a) for each quote $(t_i, p_i^b, p_i^a, q_i^b, q_i^a)$

compute the prevailing book imbalance I , as in equation (1); (b) given a quote, identify the next buy and sell trade on the tape; (c) determine the prevailing mid-price at the time of the trade; (d) compute the difference between the prevailing mid-price and the mid-price of the original quote, thus, in effect, computing a mid-to-mid return; (e) assign this return to the bucket corresponding to the original quote imbalance. A similar procedure is used for computing the waiting time until the next buy or sell trade.

Figure 2 shows average price moves and waiting times until the arrival of the next buy or sell trade, conditional on the book imbalance. The general trends of these quantities as a function of the book imbalance are similar to those obtained by using the next mid-price move as a stopping time.

However, the average price movement from the observation time to the arrival of the next sell trade displays a clear shift upwards compared with those derived from the time until the arrival of the next buy trade. This can be interpreted as the impact of the intervening buy trades between the different time spans, or the result of the information advantage of aggressive traders over those posting at the near side. As a result, traders posting their quote on the bid side of the limit order book will on average see an upwards price move by the time a sell trade matches their quote, while one posting on the ask side of the same book will experience the opposite.

Figure 3 shows the computed empirical probabilities of favourable and unfavourable price moves and of the occurrence of a matching trade, as a function of the current book imbalance, from the point of view of a broker continually ordering at the best bid or ask.

In the following sections, a three-dimensional stochastic model for the joint evolution of the bid-ask queues and the near-side trade arrival process will be used to capture these empirical features. Before developing the full model, we will first review a simpler, two-dimensional model.

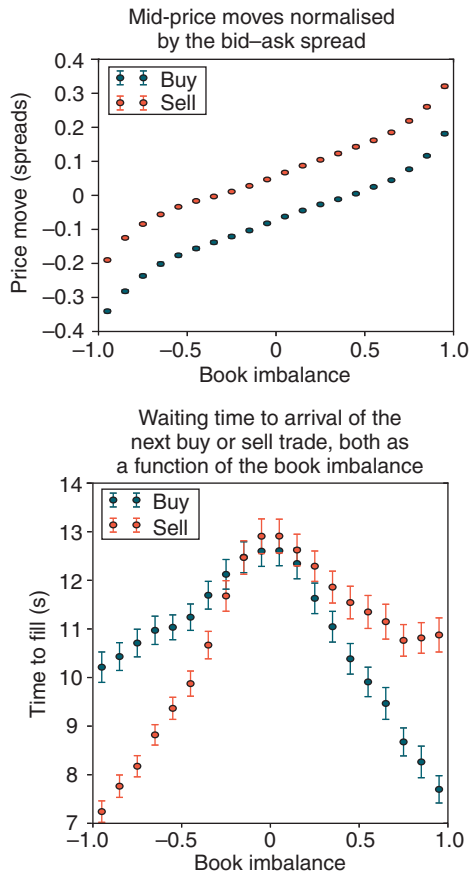
Modeling bid and ask queues

Our starting point is the two-dimensional diffusion model for the top of the book as in Cont & Larrard (2012). We model the number of shares posted at the top of the order book for the bid q^b and ask q^a as Brownian motions w^b, w^a . Because the number of shares available at the top of the book is always a positive quantity, if either Brownian motion crosses zero it is reset by drawing from two positive distributions q_{initial}^b and q_{initial}^a . One can think of these distributions as modelling the next price levels in the book; once the first level is depleted, they serve as a new starting point for the top queue. Following this interpretation, we will also assume that if the ask queue is depleted the price moves up, and if the bid is depleted the price moves down. So the pressure of incoming aggressive trades and cancellations of existing limit orders on one side makes the other side follow it, to keep the bid-ask spread fixed.

An obvious important question is whether an up-tick or a down-tick is the most likely future price move in this queue model. The probabilities can be calculated by identifying the x - and y -axes of the plane with the bid and ask quote sizes, respectively, and writing the general equation for the evolution of the hitting probability:

$$\frac{1}{2} P_{xx} + \frac{1}{2} P_{yy} + \rho_{xy} P_{xy} = 0 \quad (2)$$

2 Effect of market imbalance on price move and waiting time



Note: again, the expected price move displays a strong dependence on the book imbalance, while the waiting time shows clear dependence on the side of trade, as seen in its asymmetry

where subscripts denote differentiation with respect to that variable, and ρ_{xy} is the correlation between the depletion and replenishment of the bid and ask queues' diffusion processes, typically negative. The choice of boundary conditions selects the event with corresponding probability P ; for example, for an up-tick price movement, the boundary conditions are:

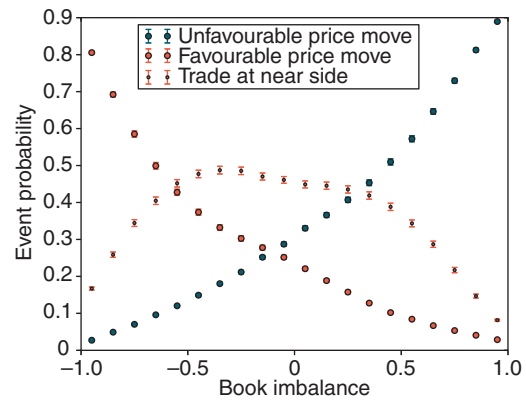
$$P(x, 0) = 1, \quad P(0, y) = 0 \quad (3)$$

Since the probability P is for event occurrence up to the first stopping time, the partial differential equation (PDE) in equation (2) is time-independent. The solution is found by taking the $T \rightarrow \infty$ limit of the corresponding finite T horizon problem.

As shown in Lipton (2001), Cont & Larrard (2012) and Lipton & Savescu (2013), among others, it is possible to solve the above PDE by introducing two changes of variables. The first transformation removes the correlation between the queue processes:

$$\begin{aligned} \alpha(x, y) &= x \\ \beta(x, y) &= \frac{(-\rho_{xy}x + y)}{\sqrt{1 - \rho_{xy}^2}} \end{aligned} \quad (4)$$

3 Market event probabilities as a function of imbalance



Note: over trading days in the first quarter of 2012. Error bars are obtained by treating averages of different trading days as independent data points

yielding the equation:

$$P_{\alpha\alpha} + P_{\beta\beta} = 0 \quad (5)$$

The second transformation casts the modified problem in polar coordinates:

$$\begin{aligned} \alpha &= r \sin(\varphi) \\ \beta &= r \cos(\varphi) \end{aligned} \quad \longleftrightarrow \quad \begin{aligned} r &= \sqrt{\alpha^2 + \beta^2} \\ \varphi &= \arctan\left(\frac{\alpha}{\beta}\right) \end{aligned} \quad (6)$$

where $\cos \varpi = -\rho_{xy}$. Then the equation for the hitting probability becomes:

$$P_{\varphi\varphi}(\varphi) = 0 \quad (7)$$

and the boundary conditions for an up-tick price movement become:

$$P(0) = 0, \quad P(\varpi) = 1 \quad (8)$$

The solution in polar coordinates is straightforward, $P(\varphi) = \varphi/\varpi$, which in the original set of coordinates has the form:

$$P(x, y) = \frac{1}{2} \left(1 - \frac{\arctan\left(\sqrt{\frac{1 + \rho_{xy}}{1 - \rho_{xy}}} \frac{y - x}{y + x}\right)}{\arctan\left(\sqrt{\frac{1 + \rho_{xy}}{1 - \rho_{xy}}}\right)} \right) \quad (9)$$

This is the probability of upward movement of a Markovian order book in the diffusive limit.

Adding trade arrival dynamics

In order to capture the joint dynamics of the bid and ask queues and trade arrival, we introduce another stochastic process ϕ to model the arrival of trades on the near side of the book, also as a Brownian motion, dw^ϕ . This does not correspond to a market observable – it is used to trigger trades when it hits zero. The queue processes q^b and q^a are driven by the addition and cancellation of limit orders in the

book, until the arrival of a trade, while the timing of near-side trade arrivals is governed by the ϕ process. The relationship between book imbalance, order flow and price variations can now be modelled by introducing correlations between dw^b , dw^a and dw^ϕ .

The introduction of the unobservable dw^ϕ leads to a non-Markovian model. As in the two-dimensional case, we assume that once a queue is depleted or a trade arrives, the process w^ϕ restarts at a value ϕ_0 which characterises the trade arrival time distribution, determined by calibration.

During the dynamic optimisation of an execution schedule, the relative likelihood of favourable and adverse price moves and of trade arrival at the near side of the book needs to be considered. In the model, these events correspond to the three-dimensional stochastic process (q^b, q^a, ϕ) crossing into the positive orthant, $x \geq 0$, $y \geq 0$, $z \geq 0$. If the corresponding hitting probability of the three-dimensional Brownian motion is $P(x, y, z)$, it satisfies:

$$\frac{1}{2}P_{xx} + \frac{1}{2}P_{yy} + \frac{1}{2}P_{zz} + \rho_{xy}P_{xy} + \rho_{xz}P_{xz} + \rho_{yz}P_{yz} = 0 \quad (10)$$

where ρ_{xy} is the correlation between the processes governing the bid and ask queues, and ρ_{xz} and ρ_{yz} are the correlation of those processes with trades at the near side. The different events are identified by boundary conditions, set to one on the plane corresponding to the event and zero elsewhere. For example, the probability of a near-side trade before any price move corresponds to the boundary conditions:

$$P(x, 0, z) = 0, \quad P(0, y, z) = 0, \quad P(x, y, 0) = 1 \quad (11)$$

As in the two-dimensional case, it is possible to remove the correlations between dw^b , dw^a and ϕ via the coordinate change:

$$\begin{aligned} \alpha(x, y, z) &= x \\ \beta(x, y, z) &= \frac{(-\rho_{xy}x + y)}{\sqrt{1 - \rho_{xy}^2}} \\ \gamma(x, y, z) &= \frac{[(\rho_{xy}\rho_{yz} - \rho_{xz})x + (\rho_{xy}\rho_{xz} - \rho_{yz})y + (1 - \rho_{xy}^2)z]}{\sqrt{1 - \rho_{xy}^2}\sqrt{1 - \rho_{xz}^2 - \rho_{yz}^2 + 2\rho_{xy}\rho_{xz}\rho_{yz}}} \end{aligned} \quad (12)$$

yielding the equation:

$$P_{\alpha\alpha} + P_{\beta\beta} + P_{\gamma\gamma} = 0 \quad (13)$$

Then a convenient set of curvilinear coordinates via the transformation:

$$\begin{aligned} \alpha &= r \sin \theta \sin \varphi & r &= \sqrt{\alpha^2 + \beta^2 + \gamma^2} \\ \beta &= r \sin \theta \cos \varphi & \theta &= \arccos(\gamma/r) \\ \gamma &= r \cos \theta & \varphi &= \arctan(\alpha/\beta) \end{aligned} \quad (14)$$

yields:

$$\frac{1}{\sin^2 \theta} P_{\varphi\varphi} + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} (\sin \theta P_\theta) = 0 \quad (15)$$

and the boundary condition for a near-side trade becomes:

$$P(0, \theta) = 0, \quad P(\varpi, \theta) = 0, \quad P(\varphi, \Theta(\varphi)) = 1 \quad (16)$$

The new integration domain after the introduction of spherical coordinates is shown in figure 4. The problem (15) can be further simplified

by introducing one extra transformation:

$$\zeta = \ln \tan \theta/2 \quad (17)$$

which changes the integration domain into the semi-infinite strip, $0 \leq \phi \leq \varpi$, $\zeta \leq Z(\phi)$, and the diffusion equation (15) into the form:

$$P_{\varphi\varphi} + P_{\zeta\zeta} = 0 \quad (18)$$

In this domain the near-side trade boundary conditions become:

$$P(0, \zeta) = 0, \quad P(\varpi, \zeta) = 0, \quad P(\varphi, Z(\varphi)) = 1 \quad (19)$$

The solution to problem (18) that satisfies the first two boundary conditions (19) can be expressed as a generalised Fourier series (Lipton 2013):

$$P(\varphi, \zeta) = \sum_{n=1}^{\infty} c_n \sin(k_n \varphi) e^{k_n \zeta} \quad (20)$$

with $k_n = \pi n / \varpi$. The expansion coefficients c_n can be determined by enforcing the third boundary condition in (19). To compute the coefficients, we introduce the integrals:

$$J_{mn} = \int_0^{\varpi} \sin(k_m \varphi) \sin(k_n \varphi) e^{(k_m + k_n) Z(\varphi)} d\varphi \quad (21)$$

and:

$$I_m = \int_0^{\varpi} \sin(k_m \varphi) e^{k_m Z(\varphi)} d\varphi \quad (22)$$

Then the third boundary condition in (19) becomes the matrix equation:

$$\sum_n J_{mn} c_n = I_m \quad (23)$$

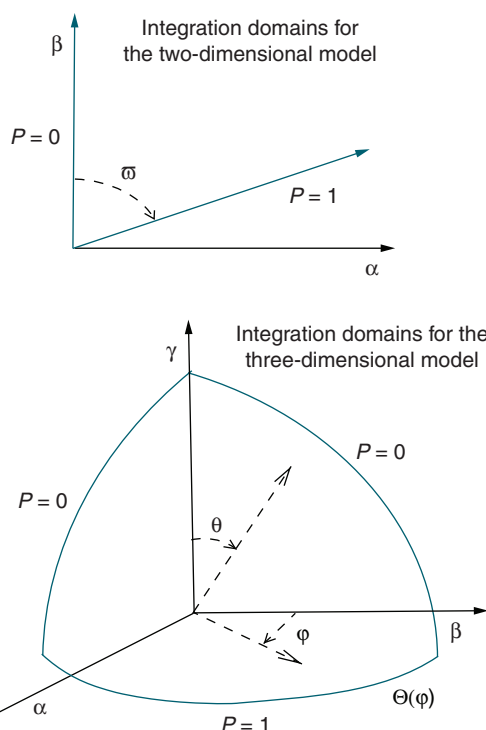
and the coefficients c_n can be computed by matrix inversion as $c = J^{-1} I$. For fixed ϕ and $\zeta < Z(\phi)$, the terms of the series (20) decay exponentially fast, more specifically, as $O(e^{-k_n(Z(\phi) - \zeta)})$, so that very few terms are needed, typically about a dozen.

When the boundary $\zeta = Z(\varphi)$ is approximately linear, the integrals I_m and J_{mn} can be computed analytically, while for a curved boundary they can be computed numerically by using a piecewise linear approximation. Figure 5 shows the solution profile for the case where the near-side trade arrival process is uncorrelated with the bid and ask queue sizes: that is, $\rho_{xz} = \rho_{yz} = 0$. In this case, the boundary is linear. It clearly demonstrates rapid convergence of the Fourier series (20) away from the boundary. All this gives a semi-analytic method of computing the probability $P(\varphi, \theta)$ or equivalently $P(x, y, z)$ for given initial values x, y, z and corresponding correlations. Both analytical and numerical methods are very fast; however, the numerical method is N times slower than the analytical one, where N is the number of piecewise linear segments approximating the boundary.

Calibration

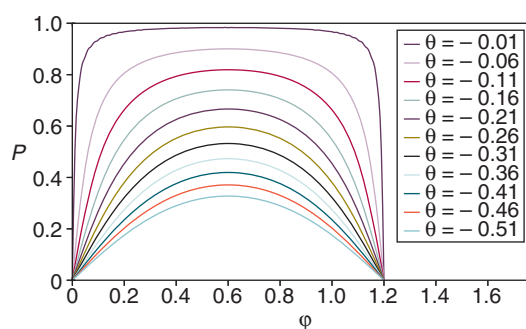
The model is calibrated jointly to the average mid-price moves shown in figure 2 and the empirical probabilities of price moves and trade arrivals summarised in figure 3. The previous section showed that

4 Integration domains for two- and three-dimensional models



Note: the boundary conditions shown correspond to an up-tick price movement in the two-dimensional case, and to a near-side trade in the three-dimensional case

5 PDE solution

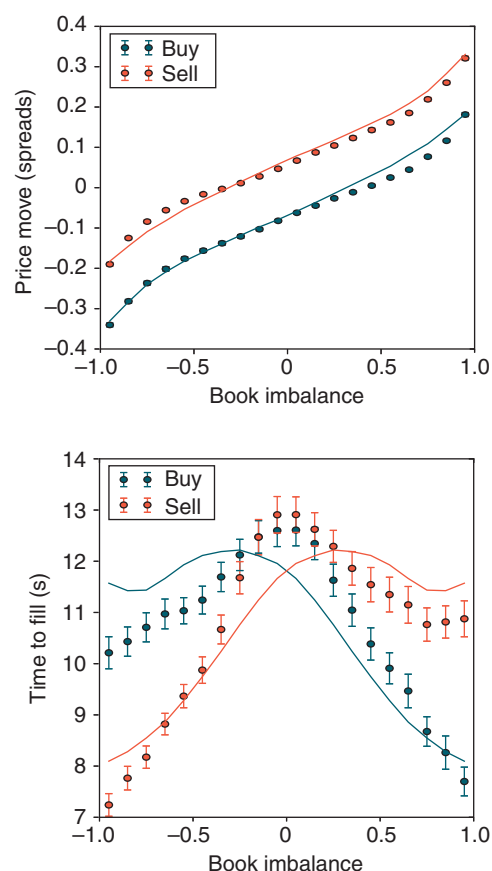


Note: solution to the problem (15) in the φ, θ space, for $\varpi = 1.2$ and boundary conditions as in (16)

the model probabilities are determined by the book imbalance, the initial position of the trade arrival process ϕ_0 and the three independent entries of the correlation matrix ρ . The book imbalance is directly observable, but the others have to be estimated. Expectations of price changes up to the stopping time defined by the arrival of a trade cannot be easily computed with analytic methods, so Monte Carlo simulations are used instead.

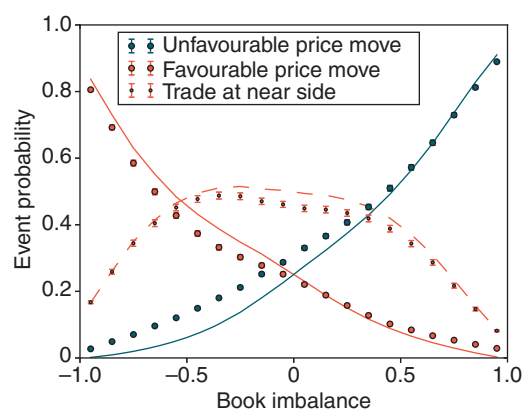
We calibrate the model to empirical probabilities of price movements for Vodafone stock, for all trading days in the first quarter of 2012. The results are shown in figures 6 and 7. The model reproduces

6 Model fit



Note: empirical and model-derived average mid-price moves normalised by the bid-ask spread and trade arrival times, as a function of book imbalance. The calibrated correlation values are $\rho_{xz} = -\rho_{yz} = 0.8$, $\rho_{xy} = -0.1$ and $\phi_0 = 3.5 \text{ sec}^{1/2}$

7 Empirical and calibrated model probabilities for the occurrence of a market event



Note: empirical and calibrated model probabilities for the first occurrence of a market event. The probability of obtaining a passive fill at the current near side is shown by a dashed red line. The probability that the price will move in favour of the broker while waiting for the fill is shown by a solid red line and the probability that the price will move against the broker by a solid blue line

the gap between the average mid-price moves until the arrival of a buy or sell trade. The gap is controlled by the correlation matrix ρ and, in particular, by the correlations between the trade arrival process and the bid and ask queue processes. This shift in average price moves conditional on the side of the trade can be interpreted as the expected slippage of a passive fill.

When executing an order, a broker will typically post a fraction of the total quantity at the near side in an attempt to save part of the spread. However, when measuring the differences in the average price achieved by aggressive and passive fills, the passive fills rarely achieve their expected savings of one full spread.

While part of this effect is generally attributed to adverse selection – other market participants taking advantage of the timing given up by the broker – it is also due to the interaction between queue depletion and trade flow, as our model predicts. In the calibration described here, it is responsible for the loss of about 60% of the theoretical spread captured by a passive fill.

Figure 6 also compares the empirical and model-derived average arrival times of a buy or sell trade as a function of book imbalance. The model captures the gross features of the empirical shapes for moderate book imbalance values, but does not reproduce the steep decrease of the arrival times at extreme book imbalance values.

Finally, the model is able to reproduce the empirical shapes of the event probabilities. As shown in figure 7, the probability of an unfavourable price movement increases as the book gets heavier towards the near side, reaching almost 90% in cases of high imbalance. Knowing this, a broker can decide to keep the order posted on the near side for moderate imbalance values and cross the spread in a highly imbalanced book. Future work should consider how to derive such optimal spread-crossing policies.

Conclusion

The microstructure of trade arrival affects the state of a limit order book. Empirically, the arrival time of trades at the near side and the dynamics of the mid-price until the arrival of a trade of a given side depend strongly on the order book imbalance. A stochastic model with correlation between the processes for the order queues at the top of the book and a process representing the arrival of the trades at the near side of the book allows the probabilities of price movement and trade arrival to be computed in a semi-analytical form. This allows for efficient calibration of the model parameters to empirical probabilities.

This model captures the dependence of trade arrival on order book imbalance, and can therefore be used to construct short-term optimal execution strategies in algorithmic trading. Such optimisations work well in normal market regimes, where book imbalance predicts trade arrival. But this predictability may be weakened by microstructure factors such as latent liquidity, or high-frequency trading, which tends to generate excessive posting and cancellation of limit orders without triggering as many trades. In these cases the calibration process will result in low correlation values between the order size and trade arrival diffusions. Future work will report on detailed empirical studies of these effects. **R**

Alex Lipton is managing director, quantitative solutions executive at Bank of America in New York and a visiting professor of quantitative finance at University of Oxford. Umberto Pesavento is vice president in London and Michael Sotiropoulos is global head in New York of algorithmic trading quantitative research at Bank of America Merrill Lynch.

Email: alex.lipton@baml.com, umberto.pesavento@baml.com, michael.sotiropoulos@baml.com

REFERENCES

Almgren R, C Thum, HL Hauptmann and H Li, 2005
Equity market impact
Risk July

Avellaneda M and S Stoikov, 2008
High-frequency trading in a limit order book
Quantitative Finance 8, pages 217–224

Bacry E and JF Muzy, 2013
Hawkes model for price and trades high-frequency dynamics
SIAM Journal of Financial Mathematics (submitted)

Bouchaud J-P, JD Farmer and F Lillo, 2008
How markets slowly digest changes in supply and demand

In T Hens and K Schenk-Hoppe, editors, *Handbook of financial markets: dynamics and evolution*
Elsevier

Bouchaud J-P, D Mezard and M Potters, 2002
Statistical properties of stock order books: empirical results and models
Quantitative Finance 2, pages 251–256

Cont R and A de Larrard, 2012
Order book dynamics in liquid markets: limit theorems and diffusion approximations
Working paper

Engle RF, 2000
The econometrics of ultra-high frequency data
Econometrica 68, pages 1–22

Hasbrouck J, 1991
Measuring the information content of stock trades
Journal of Finance 46, pages 179–207

Lipton A, 2001
Mathematical methods for foreign exchange: a financial engineer's approach
World Scientific Publishing Company

Lipton A, 2013
A structural approach to pricing credit derivatives with counterparty adjustments: additional thoughts

Global Derivatives, Trading and Risk Management, Amsterdam

Lipton A and I Savescu, 2013
CDSs, CVA and DVA – a structural approach
Risk April

Stoikov S, R Cont and R Talreja, 2010
A stochastic model for order book dynamics
Operations Research 58(3), pages 549–563

Smith E, JD Farmer, L Gillemot and S Krishnamurthy, 2003
Statistical theory of the continuous double auction
Quantitative Finance 3, pages 481–514

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.