Trading Strategies via Book Imbalance

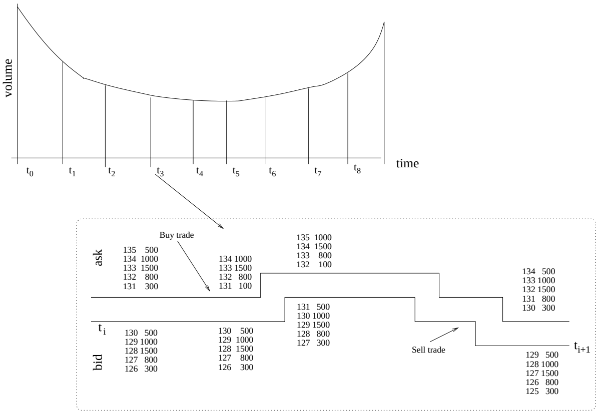
Umberto Pesavento
joint work with Alexander Lipton and Michael G. Sotiropoulos

Algorithmic Trading Quantitative Research
Bank of America Merrill Lynch

Financial Engineering Workshops, Cass Business School
City University London, 8 October 2014

Contents

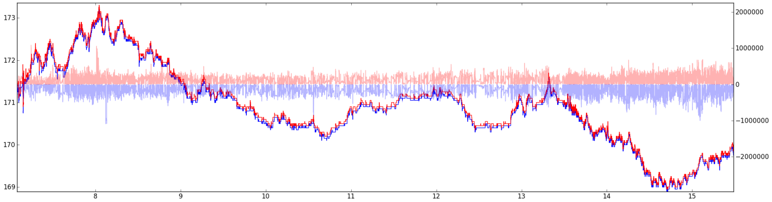
Limit order books and algorithmic trading: a top down approach



A divide and conquer approach based on 3 phases:

- calculating a time-volume schedule;
- limit order placement: optimal trading of the allocated shares within the given horizon;
- venue allocation.

Empirical observations: stopping times and book averaging

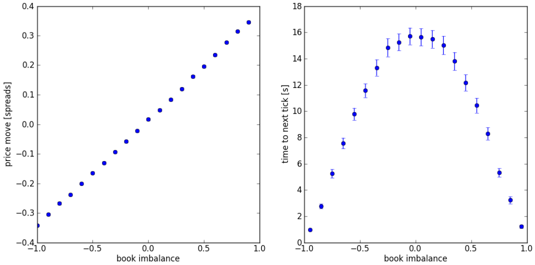


A separation of time scales:

- queue length updates ($\approx$ 3 s);
- best bid-ask updates ($\approx$ 8 s);
- trade arrivals ($\approx$ 12 s).

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
$\tilde{t}_2$	$\tilde{p}_2$	$\tilde{q}_2$	$\tilde{s}_2$	
...	...	...	...	...

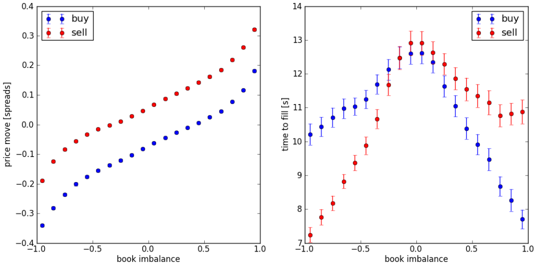
Empirical observations: mid price movements conditional on an book imbalance



Non-martingale properties of prices at small time scale:

- future price variations can be predicted by using the imbalance in the bid and ask queues of the order book $I = (q^b - q^a)/(q^b + q^a)$;
- statistically significant (about 10^7data points for the plot above);
- the effect is not large enough to lead to a straightforward arbitrage but significant enough to yield savings in execution costs.

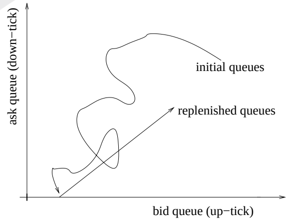
Empirical observations: trade arrivals and related stopping times



Changing the stopping time:

- the overall trend in the expected price movements as a function of the book imbalance is the same;
- conditioning on the arrival of trades on a particular side of the book breaks the symmetry in the expected waiting time and price movements.

Modeling the bid and ask queues: replenishment processes and the constant spread approximation



ask	135	3700
	133	6500
	132	9000
	131	7200
bid	130	8600
	129	5000
	128	3000
	127	4500

depletion →

135	3700
133	6500
132	9000
130	8600
129	5000
128	3000
127	4500

replenishment →

137	1500
135	3700
133	6500
132	9000
131	7200
130	8600
129	5000
128	3000

Simple models for queues and price dynamics: .

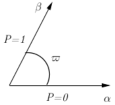
- two correlated diffusion processes to represent the bid and ask queues $(q^b, q^a) = (W^b, W^a)$;
- price moves are associated with queues depletion;
- queues replenishment, drawing from a stationary distribution;
- assume constant spreads (not a bad approximation for liquid stocks)

Modeling the bid and ask queues: time dependent dynamics

$$P_t + \frac{1}{2} P_{xx} + \frac{1}{2} P_{yy} + \rho_{xy}\, P_{xy} = 0, \tag{1}$$

where ρ_{xy} is the correlation between the processes governing the depletion and replenishment of the bid and ask queues, which typically takes a negative value in a normal market.

$$\begin{cases} \alpha(x, y) = x \\ \beta(x, y) = -\dfrac{(\rho_{xy} x - y)}{\sqrt{1 - \rho_{xy}^2}}, \end{cases} \tag{2}$$



yielding equation:

$$P_t + \frac{1}{2} P_{\alpha\alpha} + \frac{1}{2} P_{\beta\beta} = 0. \tag{3}$$

And the second to cast the problem in polar coordinates:

$$\begin{cases} \alpha = -\, r \sin(\varphi - \varpi) \\ \beta = r \cos(\varphi - \varpi) \end{cases} \longleftrightarrow \begin{cases} r = \sqrt{\alpha^2 + \beta^2} \\ \varphi = \varpi + \arctan\left(-\dfrac{\alpha}{\beta}\right) \end{cases}. \tag{4}$$

where $\cos \varpi = -\rho_{xy}$, so to yield the following equation for the hitting probabilities:

$$P_t + \frac{1}{2}\left(V_{rr} + \frac{1}{r} P_r + \frac{1}{r^2} P_{\varphi\varphi}\right) = 0. \tag{5}$$

with the final condition: $P(T, T, r, \varphi) = 0$ and boundary conditions:

$$P(t, T, 0, \varphi) = 0, \quad P(t, T, \infty, \varphi) = 0, \quad P\,(t, T, r, 0) = P_0, \quad P(t, T, r, \varpi) = P_1.$$

Modeling the bid and ask queues: Green's function formulation

We seek the Green's function to equation (5) by separating its radial and angular components:

$$G(\tau, r', \varphi') = g(\tau, r') f(\varphi'), \tag{6}$$

This leads to two equations coupled by the positive constant Λ^2 :

$$g_\tau = \frac{1}{2}\left(g_{r'r'} + \frac{1}{r'} g_{r'} - \frac{\Lambda^2}{r'^2} g\right), \tag{7}$$

$$f_{\varphi'\varphi'} = -\Lambda^2 f. \tag{8}$$

The radial part is solved by:

$$g(\tau, r') = \frac{e^{-\frac{r'^2 + r_0^2}{2\tau}}}{\tau} I_\Lambda\left(\frac{r' r_0}{\tau}\right), \tag{9}$$

where $I_\Lambda(\xi)$ is the modified Bessel function of the first kind corresponding to Λ. After applying the boundary conditions on the angular part of the equation, the final formula for the Green's function is:

$$G\left(\tau, r_0, r', \varphi_0, \varphi'\right) = \frac{2e^{-\frac{r'^2 + r_0^2}{2\tau}}}{\varpi\tau} \sum_{n=1}^{\infty} I_{\nu_n}\left(\frac{r' r_0}{\tau}\right) \sin\left(\nu_n \varphi'\right) \sin\left(\nu_n \varphi_0\right). \tag{10}$$

where $\nu_n = \frac{n\pi}{\bar{\omega}}$. Finally, we integrate the equation above to obtain the hitting probability for the of an up-tick (or down-tick) conditional on the initial condition of the queue.

$$P(t, T, r_0, \varphi_0) = -\frac{1}{2} \int_t^T \int_0^\infty G_\varphi(t' - t, r, \varpi) \frac{1}{r} dr dt'. \tag{11}$$

Modeling the bid and ask queues: infinite time limit

By writing out the explicit form for the Green's function we obtain:

$$P\left(0, r_0, \phi_0\right) = \sum_{n=1}^{\infty} \left(\int_0^T \int_0^\infty \frac{e^{-\frac{r^2+r_0^2}{2t}}}{\varpi tr} I_{\nu_n}\left(\frac{rr_0}{t}\right) dtdr \right) (-1)^{n+1} \nu_n \sin\left(\nu_n \phi_0\right). \qquad (12)$$

We reverse the order of integration and evaluate the time integral using the following expression:

$$\int_0^\infty \frac{e^{-\frac{r^2+r_0^2}{2t}}}{\varpi tr} I_{\nu_n}\left(\frac{rr_0}{t}\right) dt = \frac{1}{\varpi \nu_n r \left(\sqrt{s^2-1}+s\right)^{\nu_n}} \qquad (13)$$

where $s = (r^2 + r_0^2)/2rr_0$. We can then integrate along the radial component,

$$\int_0^\infty \frac{1}{\varpi \nu_n r \left(\max\left(\frac{r}{r_0}, \frac{r_0}{r}\right)\right)^{\nu_n}} dr = \frac{1}{\varpi \nu_n r_0^{\nu_n}} \int_0^{r_0} r^{\nu_n - 1} dr + \frac{r_0^{\nu_n}}{\varpi \nu_n} \int_{r_0}^\infty r^{-\nu_n - 1} dr = \frac{2}{\varpi \nu_n^2}. \qquad (14)$$

Finally, we sum the series to obtain:

$$P\left(0, r_0, \phi_0\right) = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin\left(\frac{\pi n}{\varpi} \phi_0\right) = \frac{\phi_0}{\varpi}. \qquad (15)$$

As expected, the result depends only on the angular distance from the barrier.

Modeling the bid and ask queues: time independent formulation I

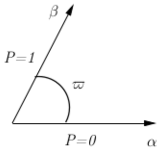
We now consider the time-independent problem from the onset:

$$\frac{1}{2}P_{xx} + \frac{1}{2}P_{yy} + \rho_{xy}\,P_{xy} = 0, \tag{16}$$

$$P\,(x,0) = 1, \quad P\,(0,y) = 0. \tag{17}$$

Again, we perform a change of coordinates to eliminate the correlation term,

$$\begin{cases} \alpha(x,y) = x \\ \beta(x,y) = \dfrac{(-\rho_{xy}x + y)}{\sqrt{1-\rho_{xy}^2}}, \end{cases} \tag{18}$$



yielding equation:

$$P_{\alpha\alpha} + P_{\beta\beta} = 0. \tag{19}$$

Modeling the bid and ask queues: time independent formulation II

We then perform a the second transformation to casts the modified problem in polar coordinates:

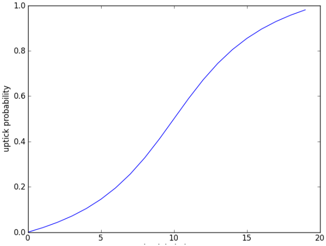
$$\begin{cases} \alpha = r\sin(\varphi) \\ \beta = r\cos(\varphi) \end{cases} \longleftrightarrow \begin{cases} r = \sqrt{\alpha^2+\beta^2} \\ \varphi = \arctan\left(\dfrac{\alpha}{\beta}\right), \end{cases} \tag{20}$$

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

$$P_{\varphi\varphi}(\varphi) = 0, \tag{21}$$

with boundary conditions $P(0) = 0$ and $P(\varpi) = 1$. In this coordinate set the solution is straightforward $P(\varphi) = \varphi/\varpi$, which in the original set of coordinates has the form:

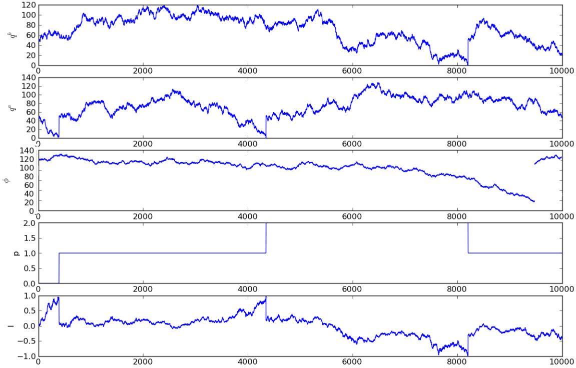
$$P(x,y) = \frac{1}{2}\left(1 - \frac{\arctan\left(\sqrt{\frac{1+\rho_{xy}}{1-\rho_{xy}}}\,\frac{y-x}{y+x}\right)}{\arctan\left(\sqrt{\frac{1+\rho_{xy}}{1-\rho_{xy}}}\right)}\right) \tag{22}$$



Adding trade arrival dynamics: the trade arrival process

In analogy with the two Brownian processes representing the bid and ask queues, we add a third (unobservable) process to model trade arrival on the near side of the book:

$$(dq^b, dq^a, d\phi) = (dw^b, dw^a, dw^\phi)$$



Adding trade arrival dynamics: handling correlation

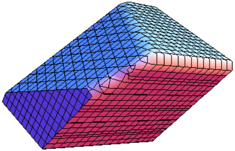
$$\frac{1}{2}P_{xx} + \frac{1}{2}P_{yy} + \frac{1}{2}P_{zz} + \rho_{xy}P_{xy} + \rho_{xz}P_{xz} + \rho_{yz}P_{yz} = 0 \qquad (23)$$

as in two dimensions, it is possible to eliminate the correlation terms,

$$\begin{cases} \alpha(x,y,z) = x \\ \beta(x,y,z) = \dfrac{(-\rho_{xy}x + y)}{\sqrt{1-\rho_{xy}^2}} \\ \gamma(x,y,z) = \dfrac{\left[(\rho_{xy}\rho_{yz} - \rho_{xz})\,x + (\rho_{xy}\rho_{xz} - \rho_{yz})\,y + (1-\rho_{xy}^2)z\right]}{\sqrt{1-\rho_{xy}^2}\sqrt{1-\rho_{xy}^2-\rho_{xz}^2-\rho_{yz}^2+2\rho_{xy}\rho_{xz}\rho_{yz}}}, \end{cases} \qquad (24)$$

to obtain:

$$P_{\alpha\alpha} + P_{\beta\beta} + P_{\gamma\gamma} = 0, \qquad (25)$$

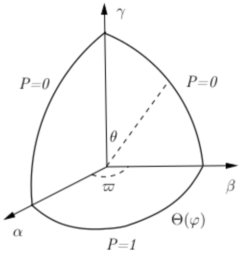


Adding trade arrival dynamics: changing the domain

Again, we can write the exit probability problem in a simpler form by changing the computational domain Ω:

$$\frac{1}{\sin^2\theta} P_{\phi\phi}(\phi,\theta) + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta P_\theta(\phi,\theta)\right) = 0, \tag{26}$$

$$P(0,\theta) = 0, \qquad P(\varpi,\theta) = 0, \qquad P(\phi,\Theta(\phi)) = 1. \tag{27}$$



$$\begin{cases} \alpha = r\sin\theta\sin\varphi \\ \beta = r\sin\theta\cos\varphi \\ \gamma = r\cos\theta \end{cases}$$

Adding trade arrival dynamics: semi analytical solutions I

We introduce a new variable $\zeta = \ln \tan \theta/2$ and rewrite the exit problem again as [Lipton 2013]:

$$P_{\phi\phi}(\phi, \zeta) + P_{\zeta\zeta}(\phi, \zeta) = 0, \tag{28}$$

computational domain is now a semi-infinite strip with curvilinear boundary

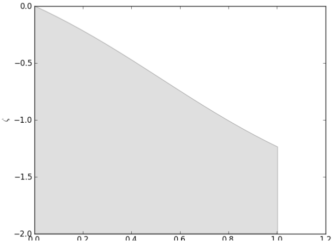
$$\zeta = Z(\phi) = \ln\left(\tan\left(\frac{\Theta(\phi)}{2}\right)\right). \tag{29}$$

We look for the solution of the Dirichlet problem for the Laplace equation in the form

$$P(\varphi, \zeta) = \sum_{n=1}^{\infty} c_n \sin(k_n \varphi), \qquad k_n = \frac{\pi n}{\varpi} \tag{30}$$

where the values of expansion coefficients c_n can be determined by enforcing the boundary condition

$$P(\varphi, \Theta(\varphi)) = 1 \tag{31}$$



$$\begin{cases} \zeta = \ln \tan \theta/2 \\ \varphi = \varphi \end{cases}$$

Adding trade arrival dynamics: semi analytical solutions II

In order to compute the coefficients, we introduce the integrals

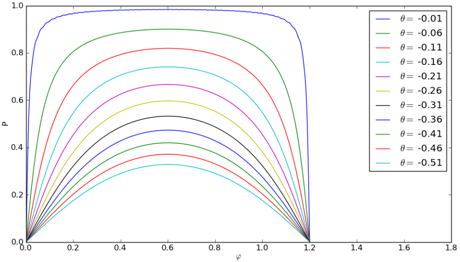
$$J_{mn} = \int_0^{\varpi} \sin(k_m\varphi)\sin(k_n\varphi)e^{(k_n+k_m)Z(\varphi)}d\varphi, \tag{32}$$

$$I_m = \int_0^{\varpi} \sin(k_m\varphi)e^{k_m Z(\varphi)}d\varphi. \tag{33}$$

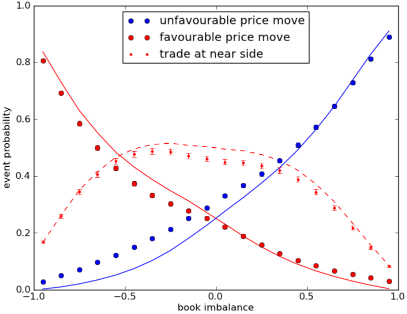
Then the boundary condition (31) becomes

$$\sum_n J_{mn}c_n = I_m, \tag{34}$$

and c_n can be computed by matrix inversion as $c = J^{-1}I$.



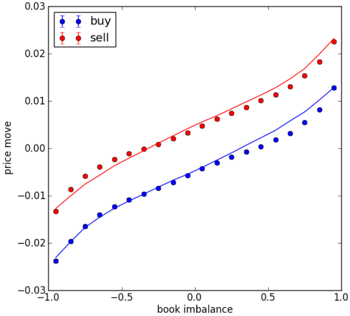
Calibration: putting everything together

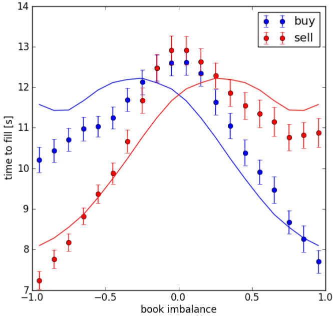


Book event probabilities as a function of the bid-ask imbalance:

- left region of the plot, price improvement is likely: get ready to reprice;
- central region of the plot, a trade on the near side is likely to anticipate an adverse price move: stay posted;
- right region of the plot: consider crossing the spread.

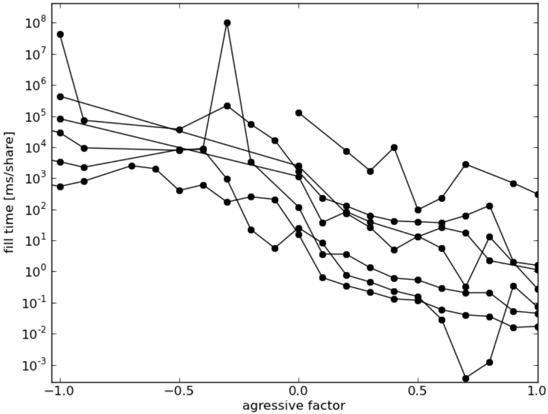
Calibration: the role of correlation





- Correlation is the main effect responsible for the symmetry breaking in the evolution of the price expectation as a function of imbalance.
- It can also explain a big part of the adverse selection effect which we observe when posting orders in a limit order book.
- The model can capture the main features of symmetry breaking in the trade arrival process.

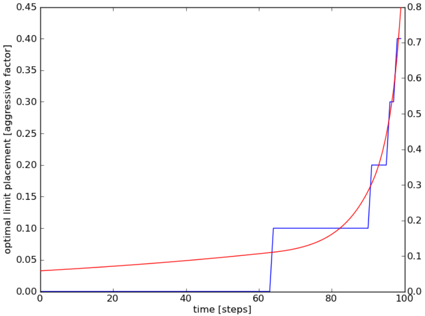
Conclusions and current work: from trade arrival rates to empirical fill probabilities



Empirical fill probabilities, learning from our own execution data:

- real data tends to be noisy, but it displays consistent trends
- parametric forms of fill probabilities as a function of the limit order placement x can be estimated, i.e. $P(x) = 1 - e^{-\beta x}$

Conclusions and current work: from empirical fill probabilities to optimization schedules, a dynamic programming approach



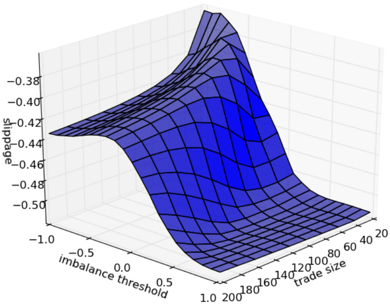
Given an approximate functional form for the fill probability, we can solve the recursive optimization problem given by:

$$E[P_i] = \min_x \left((1 - p(x))E[P_{i+1}] + p(x)x\right) \tag{35}$$

where $p(x)$ is the fill probability of a limit order with a limit price of x.

- what is the optimal placement of a limit order? (blue line)
- what is the expected fill price? (red line)

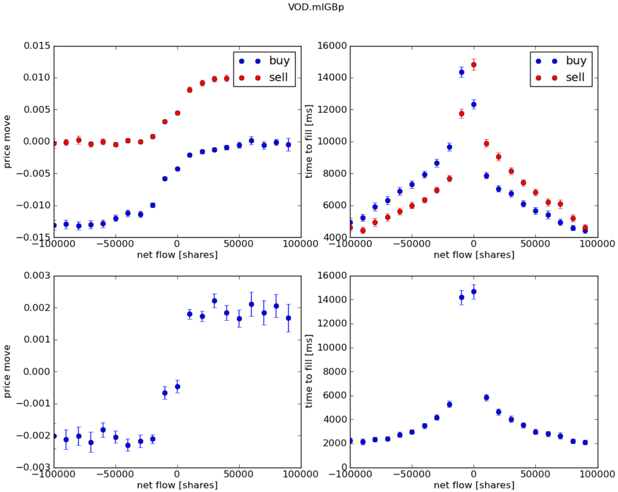
Conclusions and current work: optimizing thresholds



Going on step further, parameter selection and price slippage estimation as a function of the slice size:

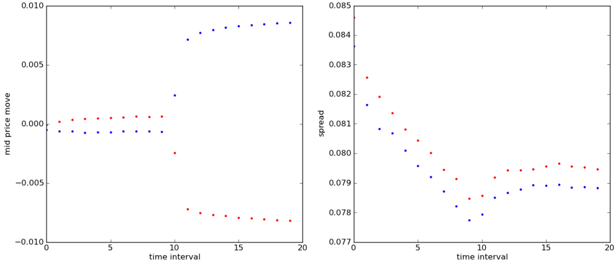
- as expected, larger slices will produce a larger slippage;
- the optimal trade off between waiting and crossing the spread depends on the size of the slice to be executed;
- an optimal ridge in the parameter space can be calculated under certain assumptions.

Appendix: order flow and impact



We can also attempt to predict price movements and arrival times by conditioning on local measurements of prevailing order flows rather than book imbalance.

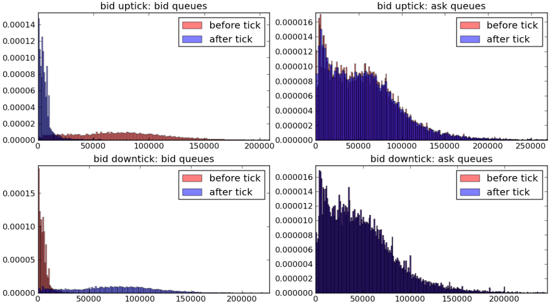
Appendix: a time dependent slice of the problem



Average time evolution of the mid-price across a trade event:

- Before trade arrival prices tends to drift towards the near side of the book;
- At trade arrival impact dominates and prices moves towards the far side of the book.

Appendix: queues depletion and replenishment



Depletion of the bid and ask queues across bid up-ticks and down-tick price movements:

- down-tick move, the initial queue size is thin while the next layer if fully formed;
- up-tick move, the previous layer is fully formed and the next queue distribution is thin;
- the ask queue is statistically unaffected.

References

[1] A. Lipton, U. Pesavento, M.Sotiropoulos, Risk, April, 2014.

[2] R. Almgren, C. Thum, H. L. Hauptmann, and H. Li. Equity market impact. Risk, 18:57, 2005.

[3] M. Avellaneda and S. Stoikov. High-frequency trading in a limit order book. Quantitative Finance, 8:217–224, 2008.

[4] J.-P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. In T. Hens and K Schenk-Hoppe, editors, Handbook of Financial Markets: Dynamics and Evolution.

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

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

[7] R. F. Engle. The econometrics of ultra-high frequency data. Econometrica, 68:1–22, 2000.

[8] J. Hasbrouck. Measuring the information content of stock trades. Journal of Finance, 46:179–207, 1991.

[9] A. Lipton and I. Savescu. CDSs, CVA and DVA - a structural approach. Risk, 26(4), 2013.

[10] S. Stoikov R. Cont and R. Talreja. A stochastic model for order book dynamics. Operations research, 58(3):549–563, 2010.

[11] E. Smith, J.D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. Quantitative Finance, 3:481– 514, 2003.