A Markov model of a limit order book: thresholds, recurrence and trading strategies

Frank Kelly, University of Cambridge
(joint work with
Elena Yudovina, University of Minnesota)

Institute for Advanced Study, City University of Hong Kong
April 2017

Outline

A *limit order book* is a trading mechanism for a single-commodity market. It is of interest

- as a model of price formation
- since it used in many financial markets.

A *limit order book* is a trading mechanism for a single-commodity market. It is of interest

- as a model of price formation
- since it used in many financial markets.

A very extensive research literature, informed by a large amount of data.

Our aim is an analytically tractable model, for a highly traded market where there may be a separation of time-scales.

A *limit order book* is a trading mechanism for a single-commodity market. It is of interest

- as a model of price formation
- since it used in many financial markets.

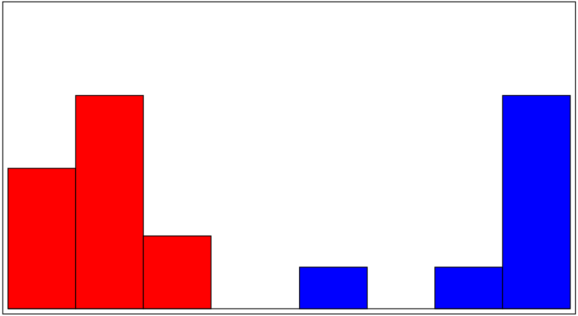
A very extensive research literature, informed by a large amount of data.

Our aim is an analytically tractable model, for a highly traded market where there may be a separation of time-scales.

Previous work (small sample!)
G. Stigler (1964), Luckock (2003)
R. Cont, S. Stoikov, and R. Talreja (2010)
X. Gao, J. G. Dai, A. B. Dieker, and S. J. Deng (2014)
P. Lakner, J. Reed, and F. Simatos (2013)
C. Maglaras, C. C. Moallemi, and H. Zheng (2014)

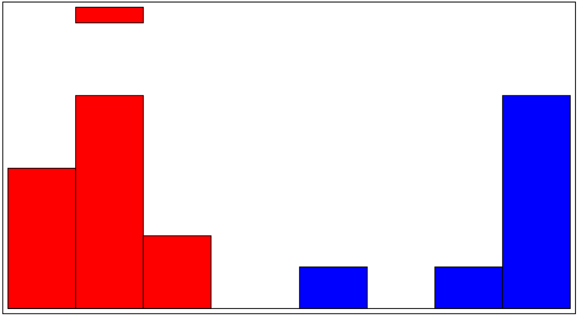
Outline

If prices are discrete, a typical system state looks like this:

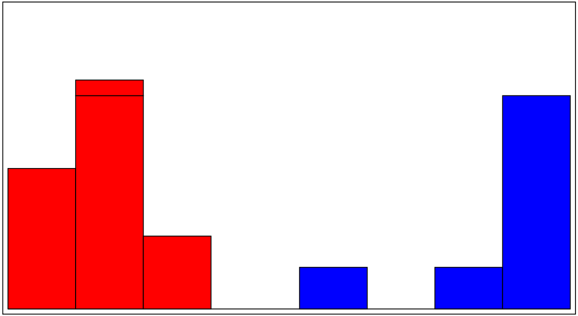


(Bids are orders to buy - red
asks are orders to sell - blue)

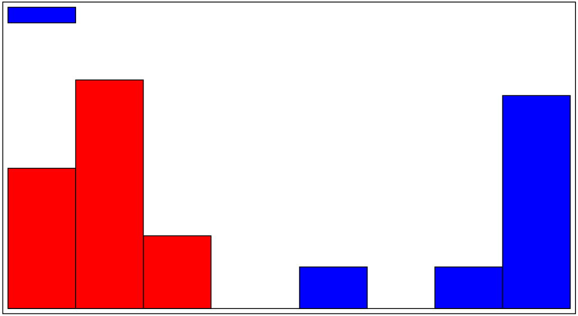
If an arriving bid is lower than all asks present ...



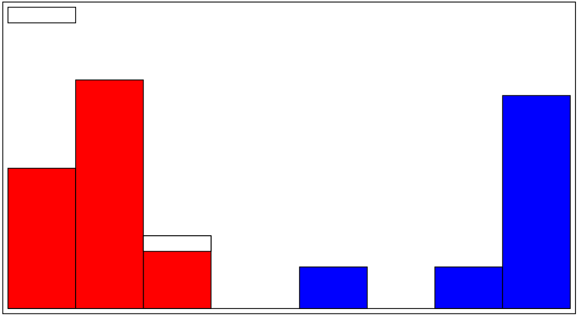
... it is added to the LOB:



If an arriving ask is lower than a bid present ...



... it is matched to the highest bid:



Assumptions:

- unit bids and unit asks arrive as independent Poisson processes of unit rate;
- the prices associated with bids, respectively asks, are independent identically distributed random variables with density $f_b(x)$, respectively $f_a(x)$.

Assumptions:

- unit bids and unit asks arrive as independent Poisson processes of unit rate;
- the prices associated with bids, respectively asks, are independent identically distributed random variables with density $f_b(x)$, respectively $f_a(x)$.

The LOB at time t is the set of bids and asks (with their prices), and our assumptions imply the LOB is a Markov process.

Assumptions:

- unit bids and unit asks arrive as independent Poisson processes of unit rate;
- the prices associated with bids, respectively asks, are independent identically distributed random variables with density $f_b(x)$, respectively $f_a(x)$.

The LOB at time t is the set of bids and asks (with their prices), and our assumptions imply the LOB is a Markov process.

Underlying idea:

- long-term investors place orders for reasons exogenous to the model, and view the market as effectively efficient;
- high-volume market with substantial trading activity even over time periods where no new information available on fundamentals of the underlying asset,

Assumptions:

- unit bids and unit asks arrive as independent Poisson processes of unit rate;
- the prices associated with bids, respectively asks, are independent identically distributed random variables with density $f_b(x)$, respectively $f_a(x)$.

The LOB at time t is the set of bids and asks (with their prices), and our assumptions imply the LOB is a Markov process.

Underlying idea:

- long-term investors place orders for reasons exogenous to the model, and view the market as effectively efficient;
- high-volume market with substantial trading activity even over time periods where no new information available on fundamentals of the underlying asset,
- leading to time-scale separation.

Assumptions:

- unit bids and unit asks arrive as independent Poisson processes of unit rate;
- the prices associated with bids, respectively asks, are independent identically distributed random variables with density $f_b(x)$, respectively $f_a(x)$.

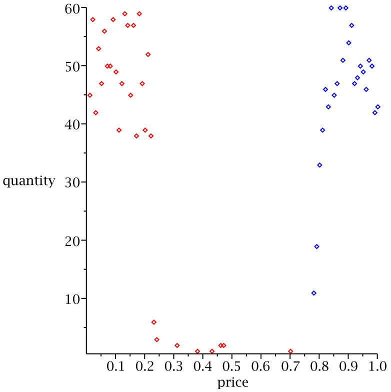
The LOB at time t is the set of bids and asks (with their prices), and our assumptions imply the LOB is a Markov process.

Underlying idea:

- long-term investors place orders for reasons exogenous to the model, and view the market as effectively efficient;
- high-volume market with substantial trading activity even over time periods where no new information available on fundamentals of the underlying asset,
- leading to time-scale separation.

We'll add high-frequency traders later.

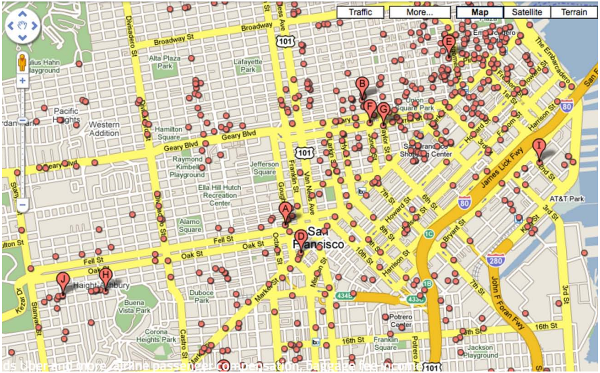
Number of bids (red) and asks (blue) at each price level, after a period (with uniform arrivals over a finite number of bins):



Other examples of two-sided queues

Early example: taxi-stand with arrivals of both taxis and travellers.

Now the queue is distributed in space with matching, and market, run by e.g. Uber.



Many other examples: Call centres, Amazon's Mechanical Turk, waiting lists for organ transplants,...

Outline

Thresholds

There exists a threshold κ_b with the following properties:

- for any $x < \kappa_b$ there is a finite time after which no arriving bids less than x are ever matched;
- and for any $x > \kappa_b$ the event that there are no bids greater than x in the LOB is recurrent.

Similarly, with directions of inequality reversed, there exists a corresponding threshold κ_a for asks.

Thresholds

There exists a threshold κ_b with the following properties:

- for any $x < \kappa_b$ there is a finite time after which no arriving bids less than x are ever matched;
- and for any $x > \kappa_b$ the event that there are no bids greater than x in the LOB is recurrent.

Similarly, with directions of inequality reversed, there exists a corresponding threshold κ_a for asks.

Intuition: eventually the highest bid and the lowest ask evolve within the interval $(\kappa_b - \epsilon, \kappa_a + \epsilon)$, for any $\epsilon > 0$.

Limiting distributions (Luckock, 2003)

There is a density $\pi_a(x)$, respectively $\pi_b(x)$, supported on (κ_b, κ_a) giving the limiting distribution of the lowest ask, respectively highest bid, in the LOB. The densities π_a, π_b solve the equations

$$f_b(x) \int_x^{\kappa_a} \pi_a(y) dy = \pi_b(x) \int_{-\infty}^x f_a(y) dy \tag{1a}$$

$$f_a(x) \int_{\kappa_b}^x \pi_b(y) dy = \pi_a(x) \int_x^{\infty} f_b(y) dy. \tag{1b}$$

Limiting distributions (Luckock, 2003)

There is a density $\pi_a(x)$, respectively $\pi_b(x)$, supported on (κ_b, κ_a) giving the limiting distribution of the lowest ask, respectively highest bid, in the LOB. The densities π_a, π_b solve the equations

$$f_b(x) \int_x^{\kappa_a} \pi_a(y) dy = \pi_b(x) \int_{-\infty}^{x} f_a(y) dy \tag{1a}$$

$$f_a(x) \int_{\kappa_b}^{x} \pi_b(y) dy = \pi_a(x) \int_x^{\infty} f_b(y) dy. \tag{1b}$$

Intuition: right-hand side of equation (1a) is the probability flux that the highest bid in the LOB is at x and that it is matched by an arriving ask with a price less than x; the left-hand side is the probability flux that the lowest ask in the LOB is more than x and that an arriving bid enters the LOB at price x; these must balance. A similar argument for the lowest ask leads to equation (1b).

Uniform example

If $f_a(x) = f_b(x) = 1, x \in (0,1)$, then
$\kappa_a = \kappa, \kappa_b = 1 - \kappa$,
$\pi_a(x) = \pi_b(1-x)$, and

$$\pi_b(x) = (1-\kappa)\left(\frac{1}{x} + \log\left(\frac{1-x}{x}\right)\right), \quad x \in (\kappa, 1-\kappa),$$

where $\kappa = w/(w+1) \approx 0.218$ with w the unique solution of $we^w = e^{-1}$.

Uniform example

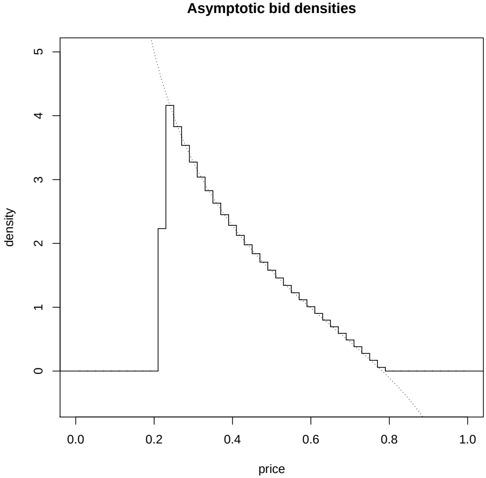
If $f_a(x) = f_b(x) = 1, x \in (0,1)$, then
$\kappa_a = \kappa, \kappa_b = 1 - \kappa$,
$\pi_a(x) = \pi_b(1-x)$, and

$$\pi_b(x) = (1-\kappa)\left(\frac{1}{x} + \log\left(\frac{1-x}{x}\right)\right), \quad x \in (\kappa, 1-\kappa),$$

where $\kappa = w/(w+1) \approx 0.218$ with w the unique solution of $we^w = e^{-1}$.

Observe that any example with $f_a = f_b$ can be reduced to this example by a monotone transformation of the price axis.

Density of the highest bid in uniform example



Binned model, with 50 bins; and continuous model.
The shoulder bin contains the (continuous) threshold.

Outline

Thresholds

- Monotonicity: if a bid is added, the future evolution of the LOB differs by either the addition of one bid or the removal of one ask; if a bid is shifted to the right, in the future evolution of the LOB the number of bids to the left of x is not increased for any x.

Thresholds

- Monotonicity: if a bid is added, the future evolution of the LOB differs by either the addition of one bid or the removal of one ask; if a bid is shifted to the right, in the future evolution of the LOB the number of bids to the left of x is not increased for any x.
- For each x, by Kolmogorov's 0–1 law,

$$\mathcal{E}^b(x) = \{\text{finitely many bids will depart from prices} \leq x\}.$$

has probability 0 or 1. Define the threshold

$$\kappa_b = \sup\{x : \mathbb{P}(\mathcal{E}^b(x)) = 1\}.$$

Similarly define the threshold κ_a using asks.

Recurrence

- Despite the existence of the thresholds κ_b, κ_a, it does not follow that the interval (κ_b, κ_a) is ever empty of both bids and asks *simultaneously*.

Recurrence

- Despite the existence of the thresholds κ_b, κ_a, it does not follow that the interval (κ_b, κ_a) is ever empty of both bids and asks *simultaneously*.
- To establish the existence of the limiting densities π_a, π_b the major challenge is to establish positive recurrence of binned models.

Recurrence

- Despite the existence of the thresholds κ_b, κ_a, it does not follow that the interval (κ_b, κ_a) is ever empty of both bids and asks *simultaneously*.
- To establish the existence of the limiting densities π_a, π_b the major challenge is to establish positive recurrence of binned models.
- After rescaling, the queue sizes and local time of the highest bid (lowest ask) in each bin converge to *fluid limits*.

Recurrence

- Despite the existence of the thresholds κ_b, κ_a, it does not follow that the interval (κ_b, κ_a) is ever empty of both bids and asks *simultaneously*.
- To establish the existence of the limiting densities π_a, π_b the major challenge is to establish positive recurrence of binned models.
- After rescaling, the queue sizes and local time of the highest bid (lowest ask) in each bin converge to *fluid limits*.
- All fluid limits tend to zero in finite time for bins inside (κ_b, κ_a). (This is the hard step: the evolution of the queues depends on which are positive rather than which are large.)

Recurrence

- Despite the existence of the thresholds κ_b, κ_a, it does not follow that the interval (κ_b, κ_a) is ever empty of both bids and asks *simultaneously*.
- To establish the existence of the limiting densities π_a, π_b the major challenge is to establish positive recurrence of binned models.
- After rescaling, the queue sizes and local time of the highest bid (lowest ask) in each bin converge to *fluid limits*.
- All fluid limits tend to zero in finite time for bins inside (κ_b, κ_a). (This is the hard step: the evolution of the queues depends on which are positive rather than which are large.)
- Deduce that the binned LOB is positive recurrent.

Recurrence

- Despite the existence of the thresholds κ_b, κ_a, it does not follow that the interval (κ_b, κ_a) is ever empty of both bids and asks *simultaneously*.
- To establish the existence of the limiting densities π_a, π_b the major challenge is to establish positive recurrence of binned models.
- After rescaling, the queue sizes and local time of the highest bid (lowest ask) in each bin converge to *fluid limits*.
- All fluid limits tend to zero in finite time for bins inside (κ_b, κ_a). (This is the hard step: the evolution of the queues depends on which are positive rather than which are large.)
- Deduce that the binned LOB is positive recurrent.
- Finally, the continuous LOB is bounded by binned models.

Theorem (ODEs)

Suppose the densities f_b and f_a on $(0,1)$ are bounded above and below. Then:

- *The highest bid and lowest ask have densities, denoted π_b and π_a; let $\varpi_b = \pi_b / f_b$ and $\varpi_a = \pi_a / f_a$.*
- *The thresholds satisfy $0 < \kappa_b < \kappa_a < 1$, $F_b(\kappa_b) = 1 - F_a(\kappa_a)$.*
- *The distribution of the highest bid is such that ϖ_b is the unique solution to the ordinary differential equation*

$$\left(-\frac{f_a(x)}{1 - F_b(x)} (F_a(x)\varpi_b(x))' \right)' = \varpi_b(x) f_b(x)$$

with initial conditions

$$(F_a(x)\varpi_b(x))|_{x=\kappa_b} = 1, \qquad (F_a(x)\varpi_b(x))'|_{x=\kappa_b} = 0$$

and the additional constraint $\varpi_b(x) \to 0$ as $x \uparrow \kappa_a$. The distribution of the lowest ask is determined by a similar ODE.

Outline

The orders we have considered so far, each with a price attached, are called *limit orders*.
Market orders request to be fulfilled immediately at the best available price. Without loss of generality assume $x \in (0, 1)$ and associate a price 1 or 0 with a market bid or market ask respectively.

The orders we have considered so far, each with a price attached, are called *limit orders*.
Market orders request to be fulfilled immediately at the best available price. Without loss of generality assume $x \in (0,1)$ and associate a price 1 or 0 with a market bid or market ask respectively.
Earlier equations (1) generalize to

$$\nu_b f_b(x) \int_x^{\kappa_a} \pi_a(y) dy = \pi_b(x) \left(\mu_a + \nu_a \int_0^x f_a(y) dy \right)$$

$$\nu_a f_a(x) \int_{\kappa_b}^x \pi_b(y) dy = \pi_a(x) \left(\nu_b \int_x^1 f_b(y) dy + \mu_b \right)$$

although now the existence of a solution is not assured.

Uniform example: stability

Let $f_a(x) = f_b(x) = 1, x \in (0, 1)$, $\nu_a = \nu_b = 1 - \lambda$ and $\mu_a = \mu_b = \lambda$. (Thus a proportion λ of orders are market orders.)

Uniform example: stability

Let $f_a(x) = f_b(x) = 1, x \in (0,1)$, $\nu_a = \nu_b = 1 - \lambda$ and $\mu_a = \mu_b = \lambda$. (Thus a proportion λ of orders are market orders.) Then, provided $\lambda < w \approx 0.278$,

$$\pi_b(\lambda; x) = \frac{1-\lambda}{1+\lambda} \cdot \pi_b\left(\frac{1+\lambda}{1-\lambda}x - \frac{\lambda}{1-\lambda}\right), \quad x \in (\kappa(\lambda), 1-\kappa(\lambda))$$

where $\pi_b(.)$ is the earlier uniform solution and

$$\kappa(\lambda) = \frac{1+\lambda}{1-\lambda} \cdot \frac{w}{1+w} - \frac{\lambda}{1-\lambda}.$$

Uniform example: stability

Let $f_a(x) = f_b(x) = 1, x \in (0, 1)$, $\nu_a = \nu_b = 1 - \lambda$ and $\mu_a = \mu_b = \lambda$. (Thus a proportion λ of orders are market orders.) Then, provided $\lambda < w \approx 0.278$,

$$\pi_b(\lambda; x) = \frac{1-\lambda}{1+\lambda} \cdot \pi_b \left(\frac{1+\lambda}{1-\lambda} x - \frac{\lambda}{1-\lambda} \right), \quad x \in (\kappa(\lambda), 1 - \kappa(\lambda))$$

where $\pi_b(.)$ is the earlier uniform solution and

$$\kappa(\lambda) = \frac{1+\lambda}{1-\lambda} \cdot \frac{w}{1+w} - \frac{\lambda}{1-\lambda}.$$

When $\lambda < w$ there is a finite (random) time after which the order book always contains limit orders of both types and no market orders of either type: hence the earlier analysis applies.

Uniform example: instability

But if $\lambda > w$ then infinitely often there will be no asks in the order book and infinitely often there will be no bids in the order book, with probability 1.

Uniform example: instability

But if $\lambda > w$ then infinitely often there will be no asks in the order book and infinitely often there will be no bids in the order book, with probability 1.
Now the difference between the number of bid and ask orders in the limit book is a simple symmetric random walk and hence null recurrent.

Uniform example: instability

But if $\lambda > w$ then infinitely often there will be no asks in the order book and infinitely often there will be no bids in the order book, with probability 1.

Now the difference between the number of bid and ask orders in the limit book is a simple symmetric random walk and hence null recurrent.

We can deduce that there will infinitely often be periods when the state of the order book contains limit orders of both types and no market orders of either type, but such states cannot be positive recurrent.

Outline

Market making

A *market maker* places an infinite number of bid, respectively ask, orders at p, respectively $q = 1 - p$, where $\kappa_b < p < q < \kappa_a$. For each pair of a bid and an ask matched, the market maker makes a profit $q - p$.

Market making

A *market maker* places an infinite number of bid, respectively ask, orders at p, respectively $q = 1 - p$, where $\kappa_b < p < q < \kappa_a$. For each pair of a bid and an ask matched, the market maker makes a profit $q - p$.

For the uniform case, the profit rate is maximized with $p \approx 0.369$, and gives a profit rate of ≈ 0.064.

Sniping

A trader immediately buys every bid that joins the LOB at price above q, and every ask that joins the LOB at price below p.

Sniping

A trader immediately buys every bid that joins the LOB at price above q, and every ask that joins the LOB at price below p. The effect on the LOB of the sniping strategy is to ensure there are no queued bids above q and no queued asks below p.

Sniping

A trader immediately buys every bid that joins the LOB at price above q, and every ask that joins the LOB at price below p.

The effect on the LOB of the sniping strategy is to ensure there are no queued bids above q and no queued asks below p.

For $p < q$, the set of bids and asks on (p, q) has the same distribution in the sniping and the market making model. For $p > q$ there are no queued orders of any kind in the interval (q, p): they are all sniped up by the trader.

Sniping

A trader immediately buys every bid that joins the LOB at price above q, and every ask that joins the LOB at price below p.
The effect on the LOB of the sniping strategy is to ensure there are no queued bids above q and no queued asks below p.
For $p < q$, the set of bids and asks on (p, q) has the same distribution in the sniping and the market making model. For $p > q$ there are no queued orders of any kind in the interval (q, p): they are all sniped up by the trader.

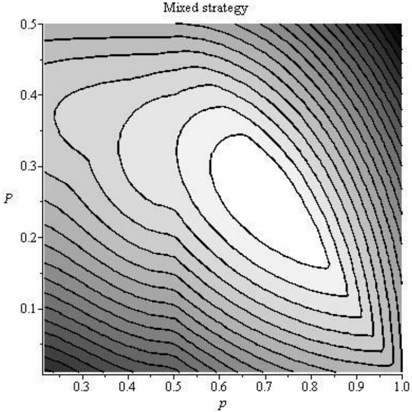
For the uniform case the profit rate is maximized at $1 - p = q = e/(e^2 + 1) \approx 0.324$ and gives a profit rate of ≈ 0.060 (lower than the optimized profit rate with a market making strategy).

A mixed strategy

Place an infinite supply of bids at P, and snipe every additional ask that land at prices $x < p$; and similarly for acquisition of bids.

A mixed strategy

Place an infinite supply of bids at P, and snipe every additional ask that land at prices $x < p$; and similarly for acquisition of bids.



For the uniform example, the optimal choice is to place an infinite bid order at $P = 1/4$, an infinite ask order at $1 - P = 3/4$ and snipe all orders that land at prices between $1/4$ and $3/4$.

Optimal profit rate $= 1/8 = 0.125$.

Competition between traders

- Between a sniper and a slower market maker: at the equilibrium of the leader-follower game, $P \approx 0.340$, $q = \sqrt{P(1-P)}$ and the profit rate of the market maker is 0.073 and of the sniper 0.020.

Competition between traders

- Between a sniper and a slower market maker: at the equilibrium of the leader-follower game, $P \approx 0.340$, $q = \sqrt{P(1-P)}$ and the profit rate of the market maker is 0.073 and of the sniper 0.020.
- Between market makers or mixed strategies: at the Nash equilibrium traders compete away the bid-ask spread and all their profits (Bertrand competition).

Competition between traders

- Between a sniper and a slower market maker: at the equilibrium of the leader-follower game, $P \approx 0.340$, $q = \sqrt{P(1-P)}$ and the profit rate of the market maker is 0.073 and of the sniper 0.020.
- Between market makers or mixed strategies: at the Nash equilibrium traders compete away the bid-ask spread and all their profits (Bertrand competition).
- Between snipers: fastest wins and monopolises profit (of rate 0.060): with frequent batch auctions, snipers must compete on price, and at the Nash equilibrium the combined profit rate is 0.042.

Conclusion

- An analytically tractable model of a limit order book on short time scales, where the dynamics are driven by stochastic fluctuations between supply and demand.
- We can use the model to analyze various high-frequency trading strategies, and the effect of competition in continuous and discrete time.
- *A Markov model of a limit order book: thresholds, recurrence, and trading strategies*
  Frank Kelly and Elena Yudovina
  Mathematics of Operations Research, to appear
  `http://arxiv.org/abs/1504.00579`