Mathematics and Financial Markets

The David Crighton Lecture 2016

Frank Kelly, University of Cambridge

Institute of Mathematics and its Applications
and the London Mathematical Society

at the Royal Society, London, 12 May 2016

Outline

STEM graduates entering Finance

The proportion of graduates working in the *Financial Activities* sector, by subject of first degree:

Degree level	Eng/Tech	Phys Sci	Maths
First degree	5.0%	8.6%	29 %
Doctorate	2.4%	2.9%	19 %

Table : First destination in finance

From *Hidden Wealth: the Contribution of Science to Service Sector Innovation*, Royal Society, 2009 (page 16).

STEM graduates entering Finance

The proportion of graduates working in the *Financial Activities* sector, by subject of first degree:

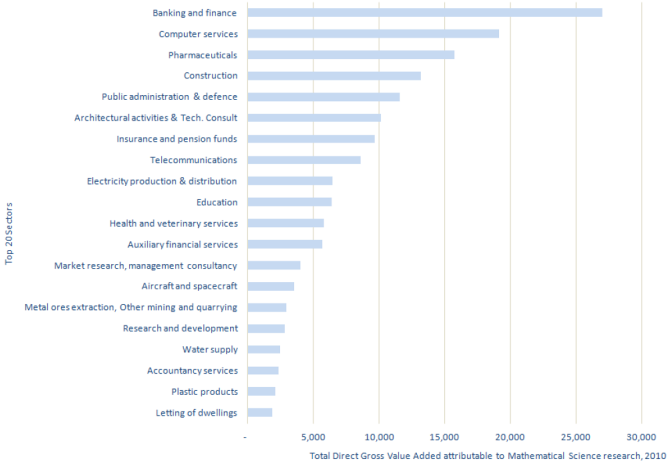
Degree level	Eng/Tech	Phys Sci	Maths
First degree	5.0%	8.6%	29 %
Doctorate	2.4%	2.9%	19 %

Table : First destination in finance

From *Hidden Wealth: the Contribution of Science to Service Sector Innovation*, Royal Society, 2009 (page 16).

A much higher proportion of graduates with degrees in mathematics were working in financial services six months after graduating, although these represent much smaller numbers of individuals than other STEM disciplines.

Figure 4.2.1: Top 20 sectors for direct mathematical science GVA in the UK, 2010, £m



Source: Deloitte using ONS data

From *Measuring the Economic Benefits of Mathematical Science Research in the UK*. Deloitte Report for the EPSRC, 2012.

FINANCIAL TIMES

"Without fear and without favour"

Saturday March 21 2009

Maths and markets

Banks need quants and geeks to recover from the crisis

Markets + maths = mayhem. That equation sums up an erroneous view of the role played by mathematics in the banking crisis, which is gaining currency in financial and regulatory circles. For example, this week's report by Lord Turner, chairman of the Financial Services Authority, blamed "misplaced reliance on sophisticated maths" for lulling banks' top managers into a false sense of security about the risks they were taking. Terms such as quant, geek and rocket scientist, once used in affectionate respect, now have darker connotations.

Mathematicians tend to be shy and retiring, compared with other professional groups, and they have not leapt up to defend themselves in public. In private, however, they are seething – understandably so, since the problem was not the maths itself but the way banks used it.

Contrary to Lord Turner's assertion, the banks' sums were not sophisticated enough. They over-simplified, and assumed away the limitations and caveats of their models. They did this to convey an illusion of accuracy and precision, and so convince the market that they had everything under control.

The standard risk measure used by the industry from the mid 1990s, known as value-at-risk or Var, was criticised by mathematicians almost from the start for the way it drew inferences about forward-looking risk from past patterns of price movements. As a result, the risk of extreme bank-shattering events was greatly underestimated.

Essentially, financial institutions told their "quants" to build mathematical models that fitted market prices – and never mind if those prices were way out of line, on any fundamental analysis. As a result, mispricing was supported by a spurious veneer of scientific respectability. And the industry was caught in a "positive feedback loop" from which no one dared walk away.

For the future we need more – and better – maths to underpin individual banks and the enhanced regulatory regime that will oversee them. Some of the expertise required is already out there, in universities, waiting to be put to use.

But financial mathematics has been underfunded, given its economic importance, and both private and public sectors must commission more research in the field. For instance, we need to know more about the way human psychology affects market models – and about the scenarios in which models break down.

At the same time, senior bankers must become better informed about the mathematical basis of their industry. Total ignorance of the "black box" trading systems used by their companies is not an acceptable excuse for failure.

Finally, mathematicians should abandon their traditional reticence and fight strongly for their discipline. Then the financial world will appreciate the true equation: markets minus maths mean mayhem.

FT editorial a response to "an erroneous view of the role played by mathematics in the banking crisis, which is gaining currency in financial and regulatory circles".

"For the future we need more - and better - maths to underpin individual banks and the enhanced regulatory regime that will oversee them."

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.

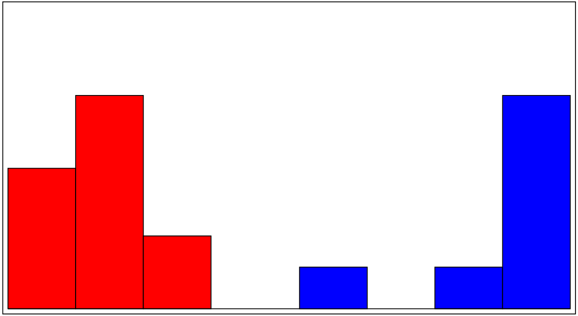
A *limit order book* is a trading mechanism for a single-commodity market. It is of interest

- 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.

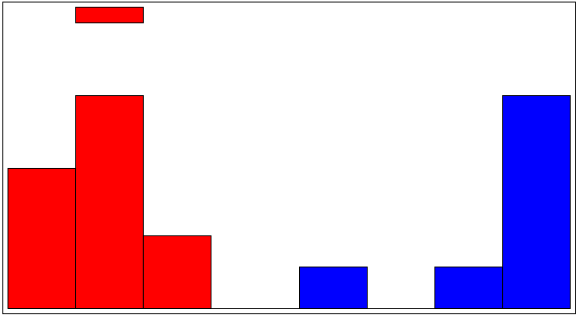
Previous work (small sample!)
R. Cont, S. Stoikov, and R. Talreja (2010)
X. Gao, J. G. Dai, A. B. Dieker, and S. J. Deng (2014)
P. Lakner, J. Reed, and F. Simatos (2013)
C. Maglaras, C. C. Moallemi, and H. Zheng (2014)

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

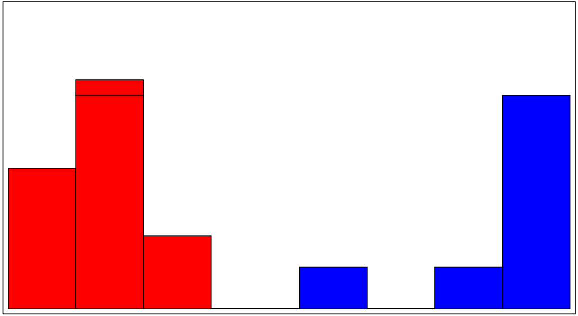


(Bids are orders to buy - red;
asks are orders to sell - blue)

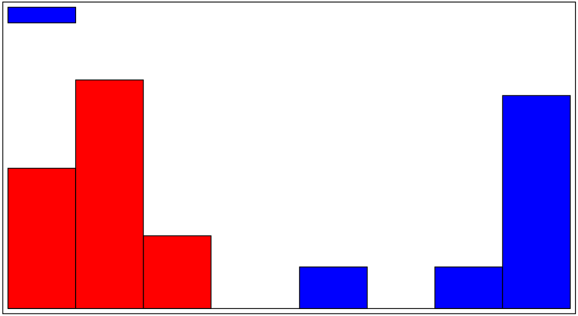
If an arriving bid is lower than all asks present ...



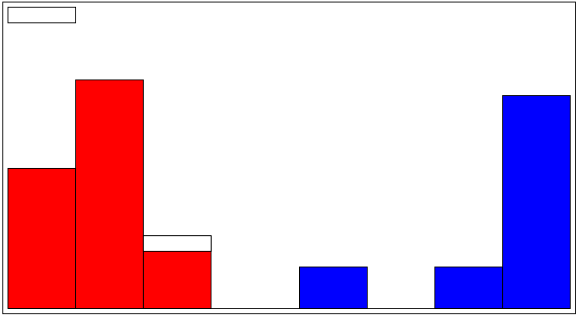
... it is added to the LOB:



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



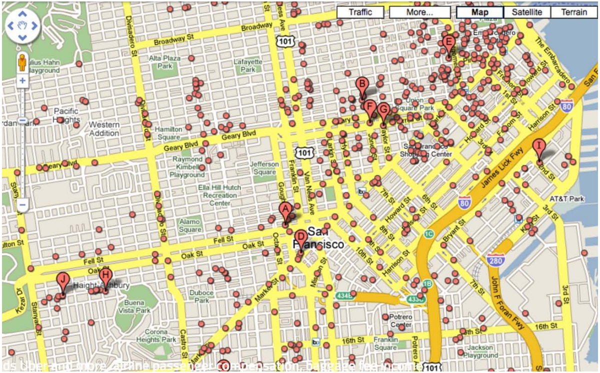
... it is matched to the highest bid:



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

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 $f_a(x)$.

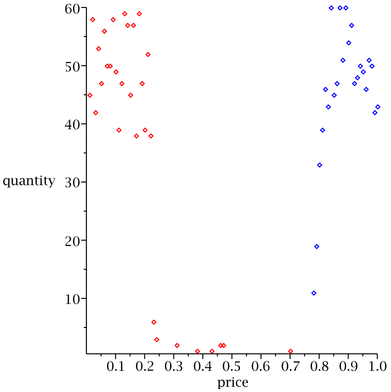
The LOB at time t is the set of bids and asks (with their prices), and our assumptions imply the LOB is a Markov process.

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,
- 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):



Outline

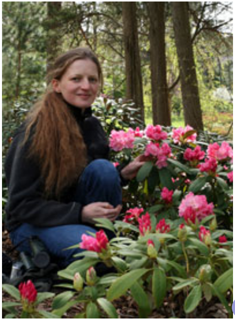
Results

The above model and the results we describe from here to the end are from:

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

http://arxiv.org/abs/1504.00579

FK and Elena Yudovina



▸ More

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

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

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.

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.
- 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*.
- To establish the existence of the limiting densities π_a, π_b we need 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.

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.

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 $p \approx 0.377$, and gives a profit rate of ≈ 0.054.

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.

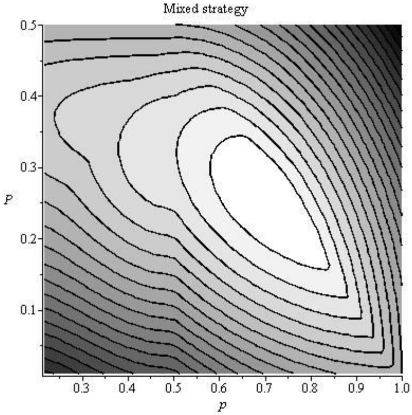
For the uniform case the profit rate is maximized at $1 - p = q = e/(e^2 + 1) \approx 0.324$ and gives a profit rate of ≈ 0.060 (higher than the optimized profit rate with a market making strategy).

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

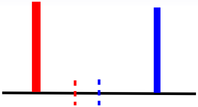
Place an infinite supply of bids at P, and snipe every additional ask that land at prices $x < p$; and similarly for acquisition of bids.



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

- Many mathematics graduates, at both first degree and doctoral level, enter the financial services sector, but not many appear to be actively engaged in the design of markets.
- Two-sided queues are becoming pervasive, as technology brings automated matching markets into more and more aspects of our lives.
- Some preliminary work on a simplified and tractable model we can use to analyze high-frequency trading strategies and competition between them:

  A Markov model of a limit order book:
  thresholds, recurrence, and trading strategies
  FK and Elena Yudovina
  `http://arxiv.org/abs/1504.00579`