

# MTH 9879 Market Microstructure Models: Spring 2019

## Lecture 1: Market mechanisms and zero intelligence models of the order book

Jim Gatheral

Department of Mathematics



### Outline of Lecture 1

- Overview of the course
- Market structure
  - Order types
  - Market types
  - Trading venues
- A brief introduction to R
- Zero intelligence models

### Why study market microstructure?

- To understand better how to trade: Specifically, how to design algorithms.
  - Market microstructure theory could be viewed as the theory underlying algorithmic trading.
- To be able to talk to informed clients.
- To understand the process of price formation.

### Aims of this course

- To be able to read and understand market microstructure literature.
  - In particular, to be familiar with well-known microstructure models.

- To be familiar with empirical characteristics of order books and time series, understanding their implications for modeling and trading.

## Philosophical overview

- Two extreme views of the market:
  - Completely random
    - The zero intelligence model associated with econophysics.
    - Use of statistical physics techniques.
  - Completely rational
    - Strategic trader models associated with efficient markets theory.
    - Game theoretic solutions.
- Reality is somewhere in-between.

## Order types

Two basic order types:

- Market orders
  - Always executed if there is sufficient quantity available.
- Limit orders
  - May be executed only if limit price reached (note "may" not "will!").
  - Typically priorities are price first, then time.

There are many variations such as:

- Fill or kill: Must be entirely filled immediately or not at all
- Immediate or cancel: Fill immediately; remaining quantity is canceled
- All or nothing: Must be entirely filled or not at all when the price is reached
- Marketable limit buy order: Buy at limit price but no higher. Remaining quantity becomes best bid.

## Hidden orders

- Users may post hidden limit orders
- Hidden orders have lower priority than visible orders at the same price level.
- Hidden orders are often placed between the visible bid and ask prices so market order submitters may sometimes get a price that is better than the visible best quote.

## Market types

- *Quote-driven* markets
    - have market makers responsible for continuously quoting two-way prices
    - For example ISE.
  - Liquidity in order-driven markets is generated by customer order flow.
    - Limit orders in the book replace market-maker quotes.
- 
- Hybrid markets such as the NYSE are mixtures of both.
    - For example, specialists are still responsible for ensuring two-way quotes in times of stress.

## Auction types

Two basic types:

- Batch auctions or *call markets*
    - For example, opening and closing auctions on NYSE.
    - Trading takes place only when the market is called.
      - Prices are "called out" until a price is found that maximizes the quantity traded.
- 
- Continuous auction or *continuous markets*
    - For example trading during the day on NYSE.
    - Trades can take place whenever the market is open.
    - In a continuous market, the limit order book effectively defines supply and demand curves for the asset being traded.

## NYSE opening auction example

Table 1: Example: 9:30am ET: Previous closing price of \$17.00

Entry Time	Buy Limit Orders	Sell Limit Orders	Order Price	Matchable	Imbalance	Indicative Price
8:30	1,000		\$19.00	0	0	-
9:00		1,000	Market	1,000	0	\$19.00
9:05	1,000		\$20.00	1,000	0	\$20.00
9:25		1,000	\$18.00	2,000	0	\$18.00

- The match price is the price that maximizes the volume that can be executed.
- There are complicated rules for tie-breaking that vary between exchanges.

## Continuous trading

- In the continuous trading session, orders are submitted and canceled.
- The set of orders with signs, sizes and prices is known as the *limit order book*.

## Trading venues

In the US, trading takes place on

- Organized exchanges such as NYSE and Nasdaq.
  - Trading can be electronic via an ECN.
  - Prices are formed through an auction process.
- Crossing networks such as LiquidNet and POSIT
  - Crosses occur with a *derivative* pricing rule.
- Dark pools with derivative pricing rules
  - Banks' internal crossing networks such as SIGMA X, CrossFinder, Instinct X
  - Other dark pools such as BIDS, Level etc.

This excerpt taken from the ITG 10-K filed Mar 1, 2007.

### POSIT

ITG POSIT crossing destinations, including POSIT Match, POSIT Now, and BLOCKalert, give buyers and sellers opportunities to match equity orders with complete confidentiality, no market impact, and the cost savings of midpoint pricing. POSIT offers unique value for traders with active, quantitative, and passive trading styles. POSIT provides access to rich, diverse liquidity, is useful for all trading styles and is especially valuable for trading small, illiquid names. All POSIT products cross at the midpoint of the National Best Bid or Offer (“NBBO”).

## Reg. NMS

From Wikipedia:

In 2005, the rules promoting the national market system were consolidated into REG NMS. Some of the more notable rules include:

Order Protection (or Trade Through) Rule - provides intermarket price priority for quotations that are immediately and automatically accessible (Rule 610)

Access Rule - addresses access to market data such as quotations (Rule 611)

Sub-Penny Rule - establishes minimum pricing increments (Rule 612)

Market Data Rules.

- This means in particular that even if there is insufficient quantity at the best quote in a given market, your market order may still be filled after rerouting to another market.

## The BATS order book

- We proceed to give a few snapshots of the BATS order book.
  - BATS is the third largest US exchange after NYSE and Nasdaq. It has the reputation of being the fastest exchange.
    - BATS stands for "Better Automated Trading System". It was bought by the CBOE in 2017.
  - You can see the order book live here:  
[http://markets.cboe.com/us/equities/market\\_statistics/book/SPY/](http://markets.cboe.com/us/equities/market_statistics/book/SPY/)  
([http://markets.cboe.com/us/equities/market\\_statistics/book/SPY/](http://markets.cboe.com/us/equities/market_statistics/book/SPY/)).

## BATS (BZX) order book for MSFT

Book Viewer >> <b>BZX Equities</b> <a href="#">BYX Equities</a> <a href="#">EDGX Equities</a> <a href="#">EDGA Equities</a> <a href="#">Market Quality Statistics</a>					
Shows the top bids and asks for any symbol.					
MSFT			Orders Accepted		Total Volume
MICROSOFT CORP COM			46,605		219,014
TOP OF BOOK			LAST 10 TRADES		
	Shares	Price	Time	Price	Shares
ASKS	753	93.6700	10:01:17	93.6300	100
	400	93.6600	10:01:17	93.6300	100
	400	93.6500	10:01:17	93.6300	500
	300	93.6400	10:01:17	93.6300	100
	200	<b>93.6300</b>	10:01:13	93.6400	100
BIDS	300	<b>93.6200</b>	10:01:13	93.6400	100
	300	93.6100	10:01:13	93.6400	100
	200	93.6000	10:01:11	93.6400	89
	500	93.5900	10:01:09	93.6500	100
	453	93.5800	10:01:09	93.6550	11
Last updated 10:01:22					

## BATS (BYX) order book for MSFT

Book Viewer >> <a href="#">BZX Equities</a> <b>BYX Equities</b> <a href="#">EDGX Equities</a> <a href="#">EDGA Equities</a> <a href="#">Market Quality Statistics</a>					
Shows the top bids and asks for any symbol.					
MSFT			Orders Accepted		Total Volume
MICROSOFT CORP COM			34,154		71,377
TOP OF BOOK			LAST 10 TRADES		
	Shares	Price	Time	Price	Shares
ASKS	100	93.6200	10:01:29	93.5750	100
	200	93.6100	10:01:29	93.5800	2
	300	93.6000	10:01:28	93.5950	100
	400	93.5900	10:01:20	93.6210	1
	300	93.5800	10:01:18	93.6300	100
BIDS	200	93.5400	10:01:17	93.6200	100
	100	93.5300	10:01:16	93.6300	100
	100	93.5200	10:01:13	93.6450	100
	353	93.5100	10:01:13	93.6400	100
	100	93.5000	10:01:02	93.6700	100
Last updated 10:01:33					

## BATS (BZX) order book for GOOG

Book Viewer >> <b>BZX Equities</b> <a href="#">BYX Equities</a> <a href="#">EDGX Equities</a> <a href="#">EDGA Equities</a> <a href="#">Market Quality Statistics</a>					
Shows the top bids and asks for any symbol.					
GOOG <input type="text"/>			Orders Accepted		Total Volume
ALPHABET INC CAP STK CL C			35,670		6,154
TOP OF BOOK			LAST 10 TRADES		
	Shares	Price	Time	Price	Shares
ASKS	100	1187.3900	10:00:43	1169.5450	26
	2	1184.5000	10:00:07	1169.9600	11
	100	1175.6200	10:00:07	1169.9600	9
	100	1174.8400	10:00:07	1169.9500	11
	22	1170.3800	10:00:07	1169.9500	9
BIDS	6	1169.1300	10:00:07	1169.8900	11
	20	1168.7900	10:00:07	1169.8900	9
	1	1166.7200	10:00:07	1169.6250	11
	100	1164.6100	10:00:07	1169.6250	20
	100	1164.3300	10:00:07	1169.6250	158
Last updated 10:01:07					

## BATS (BYX) order book for GOOG

Book Viewer >> <a href="#">BZX Equities</a> <a href="#">BYX Equities</a> <a href="#">EDGX Equities</a> <a href="#">EDGA Equities</a> <a href="#">Market Quality Statistics</a>					
Shows the top bids and asks for any symbol.					
<input type="text" value="GOOG"/> <input type="button" value="Q"/>			Orders Accepted		Total Volume
goog			35,337		2,665
TOP OF BOOK			LAST 10 TRADES		
	Shares	Price	Time	Price	Shares
ASKS	100	1217.39	10:00:33	1169.31	3
	36	1173.80	10:00:33	1169.31	1
	4	1173.40	10:00:31	1169.49	79
	36	1173.20	10:00:08	1169.51	1
	100	1170.68	09:59:39	1170.09	20
BIDS	100	1168.24	09:59:38	1170.32	14
	44	1165.70	09:59:27	1170.23	20
	44	1165.60	09:59:05	1169.70	100
	8	1165.50	09:58:12	1170.22	6
	100	1124.28	09:57:50	1170.64	8
Last updated 10:00:58					

## BATS (BZX) order book for GE

Book Viewer >> <a href="#">BZX Equities</a> <a href="#">BYX Equities</a> <a href="#">EDGX Equities</a> <a href="#">EDGA Equities</a> <a href="#">Market Quality Statistics</a>					
Shows the top bids and asks for any symbol.					
<input type="text" value="GE"/> <input type="button" value="Q"/>			Orders Accepted		Total Volume
GENERAL ELECTRIC CO COM			88,854		439,296
TOP OF BOOK			LAST 10 TRADES		
	Shares	Price	Time	Price	Shares
ASKS	3,500	16.2200	10:12:37	16.1750	100
	5,698	16.2100	10:12:32	16.1750	31
	9,700	16.2000	10:12:32	16.1800	469
	4,710	16.1900	10:12:32	16.1800	200
	5,900	16.1800	10:12:32	16.1750	100
BIDS	14,700	16.1700	10:12:32	16.1750	200
	8,300	16.1600	10:12:32	16.1750	200
	3,000	16.1500	10:12:32	16.1750	100
	3,800	16.1400	10:12:32	16.1750	100
	2,474	16.1300	10:12:32	16.1750	100
Last updated 10:14:02					



## BATS (BYX) order book for GE

Book Viewer >> <a href="#">BZX Equities</a> <b><a href="#">BYX Equities</a></b> <a href="#">EDGX Equities</a> <a href="#">EDGA Equities</a> <a href="#">Market Quality Statistics</a>					
Shows the top bids and asks for any symbol.					
GE <input type="text"/>			Orders Accepted		Total Volume
GENERAL ELECTRIC CO COM			19,749		497,480
TOP OF BOOK			LAST 10 TRADES		
	Shares	Price	Time	Price	Shares
ASKS	400	16.23	10:14:06	16.18	400
	1,500	16.22	10:14:05	16.18	100
	1,500	16.21	10:14:05	16.18	100
	3,000	16.20	10:14:01	16.18	100
	3,500	16.19	10:14:01	16.18	300
BIDS	2,700	16.17	10:14:01	16.18	400
	2,300	16.16	10:14:01	16.18	200
	1,800	16.15	10:14:01	16.18	200
	1,600	16.14	10:14:01	16.18	100
	1,500	16.13	10:14:01	16.18	400
Last updated 10:14:09					

## BATS (BZX) order book for FCX

Book Viewer >> <b>BZX Equities</b> <a href="#">BYX Equities</a> <a href="#">EDGX Equities</a> <a href="#">EDGA Equities</a> <a href="#">Market Quality Statistics</a>					
Shows the top bids and asks for any symbol.					
FCX			Orders Accepted		Total Volume
FREEPORT-MCMORAN INC CL B			29,571		140,418
TOP OF BOOK			LAST 10 TRADES		
	Shares	Price	Time	Price	Shares
ASKS	900	19.56	10:21:48	19.52	100
	1,616	19.55	10:21:12	19.51	100
	1,000	19.54	10:21:12	19.51	100
	1,400	19.53	10:21:08	19.51	100
	400	19.52	10:21:08	19.51	100
BIDS	900	19.51	10:21:08	19.51	100
	1,300	19.50	10:21:08	19.51	100
	1,000	19.49	10:20:57	19.52	100
	800	19.48	10:20:57	19.52	100
	700	19.47	10:20:57	19.52	100
Last updated 10:22:17					

## BATS (BYX) order book for FCX

Book Viewer >> <a href="#">BZX Equities</a> <b>BYX Equities</b> <a href="#">EDGX Equities</a> <a href="#">EDGA Equities</a> <a href="#">Market Quality Statistics</a>					
Shows the top bids and asks for any symbol.					
FCX			Orders Accepted		Total Volume
FREEPORT-MCMORAN INC CL B			16,185		72,913
TOP OF BOOK			LAST 10 TRADES		
	Shares	Price	Time	Price	Shares
ASKS	100	19.5700	10:21:57	19.5200	100
	200	19.5600	10:21:57	19.5200	100
	800	19.5500	10:21:54	19.5200	100
	500	19.5400	10:21:48	19.5200	100
	400	19.5300	10:21:43	19.5200	100
BIDS	500	19.5100	10:21:33	19.5100	100
	662	19.5000	10:21:31	19.5050	100
	200	19.4900	10:21:28	19.5100	100
	500	19.4800	10:21:23	19.5100	100
	500	19.4600	10:21:21	19.5100	100
Last updated 10:21:59					

## BZX and BYX

- BZX gives a rebate to limit order submitters
  - There is an incentive to place limit orders so the size at best quote is large.
  - Market orders are small.
  - For small tick stocks, spreads are smaller.
- BYX gives a rebate to market order submitters.
  - Queue sizes are smaller.
  - Market orders are bigger.

- Because of different fee structures on different exchanges, the displayed spread is not the true spread.
- Current BATS fee schedules are available here:  
[http://markets.cboe.com/us/equities/membership/fee\\_schedule/bzx/](http://markets.cboe.com/us/equities/membership/fee_schedule/bzx/)  
 ([http://markets.cboe.com/us/equities/membership/fee\\_schedule/bzx/](http://markets.cboe.com/us/equities/membership/fee_schedule/bzx/)) and here  
[http://markets.cboe.com/us/equities/membership/fee\\_schedule/byx/](http://markets.cboe.com/us/equities/membership/fee_schedule/byx/)  
 ([http://markets.cboe.com/us/equities/membership/fee\\_schedule/byx/](http://markets.cboe.com/us/equities/membership/fee_schedule/byx/)).

## Large and small tick stocks

- Stocks where the tick size is large relative to the price movement per trade are *large tick* stocks. Stocks where the tick size is small relative to the price movement per trade are *small tick* stocks.
  - GE and FCX are large tick stocks.
  - GOOG and MSFT are small tick stocks.

- A large tick stock has a spread of one tick nearly all of the time.
- Conversely, the spread of a small tick stock is typically greater than one tick (one cent in the US).

## The Santa Fe zero-intelligence model

In this model due to Smith, Farmer, Gillemot and Krishnamurthy:

- Limit orders can be placed at any integer price level  $p$  where  $-\infty < p < \infty$ .
  - If worried about negative prices, think of these as being logarithms of the actual price.
- Limit sell orders may be placed at any level greater than the best bid  $b(t)$  at time  $t$  and limit buy orders at any level less than the best offer  $a(t)$ .

- In particular, just as in real markets, limit orders may be placed inside the spread (if the current spread is greater than one tick).
- Market orders arrive randomly at rate  $\mu$ .
- Limit orders (per price level) arrive at rate  $\lambda$ .
- A proportion  $\nu$  of existing limit orders is canceled.
- All market orders and limit orders are for one share.
- All the order flows are modeled as Poisson processes.

## Analysis of the Santa Fe order book model

- The Santa Fe model depends on so few parameters that, even though it is capable of generating complex behavior, it remains amenable to analytical investigation.
- For example, it is easy to see that the asymptotic book depth far away from the best quote must be given by  $\lambda/\nu$ .

The argument goes as follows:

## Asymptotic book depth

- In the steady-state, orders arriving into the book must balance orders leaving the book.
- Far away from the best quote, the probability of a limit order leaving the book as the result of an execution against a market order is very small.

- Thus we need only consider new orders arriving balancing existing orders leaving due to cancelation.
  - At a given price, for a fixed time interval  $\Delta$ , the expected number of orders arriving is  $\lambda\Delta$  and the expected number of orders canceled is  $\nu d\Delta$  where  $d$  is the number of shares at the book at that price level.
  - For these to balance, we must have  $\lambda\Delta = \nu d\Delta$ , or equivalently,

$$d = \frac{\lambda}{\nu}.$$

## Complex interactions

- Although the model is simple to describe, its behavior can be rather complex.
- For example, suppose there is one share at the best bid when the spread is two ticks.
  - If a new buy order is placed inside the spread one tick below the best offer  $a_t$ , the best bid  $b_t$  increases and the spread  $s_t$  decreases to one tick.
  - If a market sell order arrives, the book will revert to its initial state.

- Now suppose that the same events occur but in the reverse order.
  - The market sell order increases the spread to three ticks and a new limit buy order is placed one tick below the best offer.
  - The spread decreases to one tick.
- The resulting book shape and even the mid-quote are quite different.

## Distribution of order depth

- In the Santa Fe model, the distribution of order depth is a function of distance to the best quote.
  - Conditional on no change in the best quotes, order depth would be i.i.d at any given level of the book deeper than the best quote.
  - In fact, as shown by SFGK, this distribution is Poisson with mean  $\lambda/\nu$ .
  - However, as the order book evolves, the best quote prices do move, and each level in the book retains some memory of having been "visited" by the market price at some prior time.

- Strictly speaking, order depth distributions are therefore conditional on the entire history of the order book and in particular on the distance to the best quote.
- The further away a given level is from the best quote, the closer the order depth distribution will be to Poisson.

## Dimensional analysis

- Dimensional analysis is a favorite tool of physicists used to guess the relationships between model outputs and parameters.
  - Effectively using information about scaling relationships.

- The dimensions of the parameters of the Santa Fe model are as follows:

Parameter	Dimensions
$\mu$	shares/T
$\lambda$	shares/(ticks T)
$\nu$	1/T

- In the following we use these parameter dimensions to guess the functional form of relationships between measurable quantities of interest and the input parameters.

## Inverse table

- For convenience, we provide (unique) expressions for dimensions in terms of input parameters.

- Inverting the previous table gives

Dimensions	Parameter combinations
shares	$\mu / \nu$
T	$1 / \nu$
ticks	$\mu / \lambda$

## Asymptotic book depth

- The asymptotic depth  $d$  of the book far away from mid-market has the dimensions of *shares per tick*. The only combination of parameters with these dimensions is  $\lambda/\nu$ .
- As we saw earlier, the exact result is in fact just  $\lambda/\nu$  so the guess in this case is exactly correct.

## The spread

- The spread has dimensions of *ticks* and the only combination of parameters with these dimensions is  $\mu/\lambda$ .
- Arguing more directly, inside the spread, limit orders are mostly removed by market orders.
  - The total rate of arrival of limit orders is  $2 \lambda s$  where  $s$  is the steady-state spread because as mentioned earlier, buy and sell orders may be placed inside the spread.
  - The rate of removal of limit orders is just the rate of market orders  $\mu$ .

- We deduce that (roughly at least)

$$s \sim \frac{\mu}{2 \lambda}$$

## The slope of the order book

- This quantity has dimensions of *shares per tick per tick* (the derivative with respect to ticks of the limit order density).
- The only combination of parameters that has these dimensions is

$$\frac{\lambda^2}{\mu \nu}.$$

## Volatility

- The variance of the times series of trade prices generated in this model must have dimensions  $\text{ticks}^2/T$ .
- Dimensional arguments imply that variance per unit time must therefore be proportional to

$$\frac{\mu^2 \nu}{\lambda^2}$$

- Thus volatility must scale as

$$\frac{\mu \sqrt{\nu}}{\lambda}$$

## Average quote size

- The average quote size  $Q$  should be given roughly by the cumulative steady-state inventory  $s/2$  ticks away from mid-market where  $s$  is the average spread.
- Thus

$$Q \sim s^2 \frac{\lambda^2}{\mu \nu} \sim \frac{\mu^2}{\lambda^2} \frac{\lambda^2}{\mu \nu} = \frac{\mu}{\nu}$$

as we would expect because once again,  $\mu/\nu$  is the only combination of parameters with dimensions of *shares*.

## Characteristic share quantity

The only quantity that can be formed from the model parameters with the dimensions of *shares* is the characteristic share quantity

$$N_c = \frac{\mu}{2\nu}$$

## Market impact

The most familiar market impact model is  $\sigma \sqrt{L}$  where  $\sigma$  is volatility the liquidity  $L$  is fraction of average daily volume. In this model, for a given number of shares  $n$ ,

$$L = \frac{n}{\mu}$$

and as shown above

$$\sigma \sim \frac{\mu \sqrt{\nu}}{\lambda}$$

Re-expressing that model of market impact in terms of the parameters of the Santa Fe model gives

$$I(n) \sim \frac{\mu \sqrt{\nu}}{\lambda} \sqrt{\frac{n}{\mu}} = \frac{1}{\lambda} \sqrt{\mu \nu n}$$

which has the correct dimensions of *ticks*.

## Market impact in general

- In general, no matter what the form of market impact, for a given number of shares  $n$  the impact must have dimensions of *ticks*.
- The only combination of parameters with dimensions of *ticks* is  $\mu/\lambda$ .

- We deduce that the general form of market impact must be

$$I(n) \sim \frac{\mu}{\lambda} f\left(\frac{n}{N_c}\right) = \frac{\mu}{\lambda} f\left(\frac{2\nu n}{\mu}\right)$$

for some function  $f(\cdot)$ .

- Clearly, the square-root model of market impact is consistent with this.

## Realism of Santa Fe model

- In real markets, order depth falls off rapidly away from the best quote.
- We would therefore not expect the behavior of the Santa Fe model deep in the order book to accurately mimic a real market.
- However, close to the best quote, the Santa Fe model *does* appear to capture some salient aspects of real markets.

In [ ]:

- For example:
  - A model in a similar style was shown by Bollerslev, Domowitz and Wang to accurately predict the distribution of bid-offer spreads in the foreign exchange market.
  - Farmer, Patelli and Zovko estimated  $\mu$ ,  $\lambda$  and  $\nu$  from London Stock Exchange SETS order book data, comparing average actual bid-offer spreads and volatilities against Santa Fe model predictions and finding agreement to be quite good.
- As we shall see, there are inevitably some aspects of real markets that zero-intelligence models cannot mimic, notably those associated with strategic behavior of traders.

## Implementation of the Santa Fe model

- It is impossible to simulate order arrivals and cancelations at integer price levels from  $-\infty$  to  $+\infty$ .
- So consider only order arrivals and cancelations in a moving band of width  $L$  centered around the current best quotes.
  - $L$  should be chosen conservatively so as to ensure minimal edge effects.



- Within the band, the arrival rate of limit orders is  $\lambda$ , cancelation rate is  $\nu$  times outstanding shares.
- Outside the band, orders may neither arrive nor be canceled.

## Notation

Table 2: Order book events

Event	Notation
Limit buy order	LB
Market buy order	MB
Cancel buy order	CB
Market sell order	MS
Limit sell order	LS
Cancel sell order	CS

We denote the probability of an event by  $\mathbb{P}_{event}$

## Santa Fe pseudo-code

- Compute the best bid  $B_t$  and best offer  $A_t$ .
  - Compute the number  $n_B$  of shares on the bid side of the book from level  $A_t - 1$  to level  $A_t - L$ .
  - Compute the number  $n_A$  of shares on the offered side of the book from level  $B_t + 1$  to level  $B_t + L$ .
- 
- Draw a new event according to the relative probabilities  $\{\mathbb{P}_{MB}, \mathbb{P}_{MS}, \mathbb{P}_{LB}, \mathbb{P}_{LS}, \mathbb{P}_{CS}, \mathbb{P}_{CB}\} \propto \{\mu/2, \mu/2, L\alpha, L\alpha, \delta n_A, \delta n_B\}$ .
    - If the selected event is a limit order, draw the relative price level from  $\{1, 2, \dots, L\}$ .
    - If the selected event is a cancelation, select randomly which order within the band to cancel.
  - Update the order book and increment  $t$ .

## Outputs of simulation

This simulation procedure yields a time series of order book data from which we extract quantities of interest such as spread, mid-quote etc.

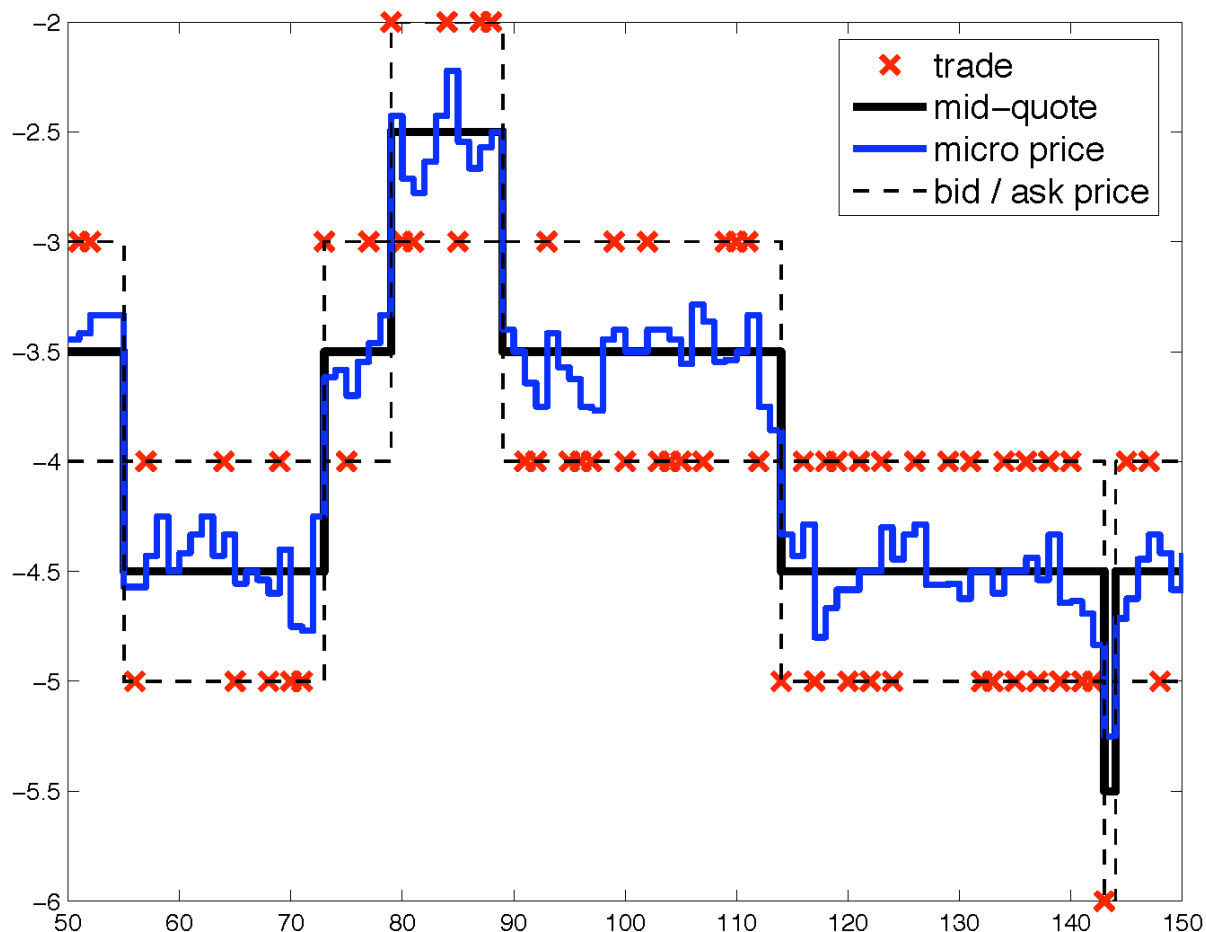


Figure 1: A typical sample path

## Introduction to R

There are many places to learn about R. I recommend:

- *R for Beginners* by Emmanuel Paradis available at [http://cran.r-project.org/doc/contrib/Paradis-rdebuts\\_en.pdf](http://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf) ([http://cran.r-project.org/doc/contrib/Paradis-rdebuts\\_en.pdf](http://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf))
- *R for MATLAB users* available at <http://mathesaurus.sourceforge.net/octave-r.html> (<http://mathesaurus.sourceforge.net/octave-r.html>)

## R implementation of ZI simulation

```
In [1]: download.file(url="https://mfe.baruch.cuny.edu/wp-content/uploads/2019/01/ziSetup.zip")
        unzip(zipfile="ziSetup.zip")

        source("ziSetup.R")
```

## What is inside ziSetup.R?

```
In [2]: #Book setup
L <- 30 #Set number of price levels to be included in iterations

# Generate initial book
LL <- 1000 #Total number of levels in buy and sell books
```

```
In [3]: # Initialize book with asymptotic depth of 5 shares
initializeBook5

function ()
{
  Price <- -LL:LL
  buySize <- c(rep(5, LL - 8), 5, 4, 4, 3, 3, 2, 2, 1, rep(0,
    LL + 1))
  sellSize <- c(rep(0, LL), 0, 1, 2, 2, 3, 3, 4, 4, 5, rep(5,
    LL - 8))
  book <- data.frame(Price, buySize, sellSize)
  if (logging == T) {
    eventLog <- as.data.frame(matrix(0, nrow = numEvents,
      ncol = 2))
    colnames(eventLog) <- c("Type", "Price")
    count <- 0
    eventType <- c("LB", "LS", "CB", "CS", "MB", "MS")
    eventDescr <- NA
  }
}
```

```
In [4]: #Various utility functions
bestOffer
```

```
function ()
{
  min(book$Price[book$sellSize > 0])
}
```

and similiarly bestBid, spread, mid

```
In [5]: #Functions to find mid-market
bidPosn
```

```
function ()
length(book$buySize[book$Price <= bestBid()])
```

and similarly askPosn, midPosn.

```
In [7]: #Display center of book  
go
```

```
function ()  
{  
  book[(midPosn() - 20):(midPosn() + 20), ]  
}
```

```
In [8]: #Display book shape  
bookShape
```

```
function (band)  
{  
  c(book$buySize[midPosn() + (-band:0)], book$sellSize[midPosn() +  
    1:band])  
}
```

```
In [9]: bookPlot
```

```
function (band)  
{  
  plot((-band:band), bookShape(band), col = "red", type = "l",  
        xlab = "Price", ylab = "Quantity")  
}
```

```
In [10]: #Choose from L whole numbers in (1,...,L) with uniform probability  
pick
```

```
function (m)  
{  
  sample(1:m, 1)  
}
```

```
In [11]: # Switch logging on  
logging <- T
```

```
In [12]: #Buy limit order
         limitBuyOrder
```

```
function (price = NA)
{
  if (is.na(price)) {
    prx <- (bestOffer() - pick(L))
  }
  else prx <- price
  if (logging == T) {
    eventLog[count, ] <- c("LB", prx)
  }
  book$buySize[book$Price == prx] <- book$buySize[book$Price ==
    prx] + 1
}
```

```
In [13]: #Sell limit order
         limitSellOrder
```

```
function (price = NA)
{
  if (is.na(price)) {
    prx <- (bestBid() + pick(L))
  }
  else prx <- price
  if (logging == T) {
    eventLog[count, ] <- c("LS", prx)
  }
  book$sellSize[book$Price == prx] <- book$sellSize[book$Price ==
    prx] + 1
}
```

```
In [14]: #Cancel buy order
         cancelBuyOrder
```

```
function (price = NA)
{
  q <- pick(nb)
  tmp <- cumsum(rev(book$buySize))
  posn <- length(tmp[tmp >= q])
  prx <- book$Price[posn]
  if (!is.na(price)) {
    prx <- price
  }
  if (logging == T) {
    eventLog[count, ] <- c("CB", prx)
  }
  book$buySize[posn] <- book$buySize[posn] - 1
}
```

```
In [15]: #Cancel sell order
cancelSellOrder
```

```
function (price = NA)
{
  q <- pick(ns)
  tmp <- cumsum(book$sellSize)
  posn <- length(tmp[tmp < q]) + 1
  prx <- book$Price[posn]
  if (!is.na(price)) {
    prx <- price
  }
  if (logging == T) {
    eventLog[count, ] <- c("CS", prx)
  }
  book$sellSize[posn] <- book$sellSize[posn] - 1
}
```

```
In [16]: #Market buy order
marketBuyOrder
```

```
function ()
{
  prx <- bestOffer()
  if (logging == T) {
    eventLog[count, ] <- c("MB", prx)
  }
  book$sellSize[book$Price == prx] <- book$sellSize[book$Price ==
    prx] - 1
}
```

```
In [17]: #Market sell order
marketSellOrder
```

```
function ()
{
  prx <- bestBid()
  if (logging == T) {
    eventLog[count, ] <- c("MS", prx)
  }
  book$buySize[book$Price == prx] <- book$buySize[book$Price ==
    prx] - 1
}
```

## Generate an event and update the buy and sell books

- Note that limit orders may be placed inside the spread

In [18]: generateEvent

```
function ()
{
  nb <- sum(book$buySize[book$Price >= (bestOffer() - L)])
  ns <- sum(book$sellSize[book$Price <= (bestBid() + L)])
  eventRate <- nb * nu + ns * nu + mu + 2 * L * lambda
  probEvent <- c(L * lambda, L * lambda, nb * nu, ns * nu,
    mu/2, mu/2)/eventRate
  m <- sample(1:6, 1, replace = TRUE, probEvent)
  switch(m, limitBuyOrder(), limitSellOrder(), cancelBuyOrder(),
    cancelSellOrder(), marketBuyOrder(), marketSellOrder())
}
```

## Average book shape in the ZI simulation

In [19]: *# Figure 2: Average book shape*

```
logging <- F # Very important for speed!

lambda <- 1
mu <- 50
nu <- 1/5
initializeBook5()
# Burn in for 100 events
for(count in 1:100){
  generateEvent()
}
numEvents <- 10000 # Average over 10,000 events
avgBookShape <- bookShape(20)/numEvents
for(count in 2:numEvents){
  generateEvent()
  avgBookShape <- avgBookShape+bookShape(20)/numEvents
}
```

In [20]: library(repr)  
options(repr.plot.height=5)

```
In [21]: plot(-20:20, avgBookShape, main=NA, xlab="Relative price", ylab="# Shares", col=
```

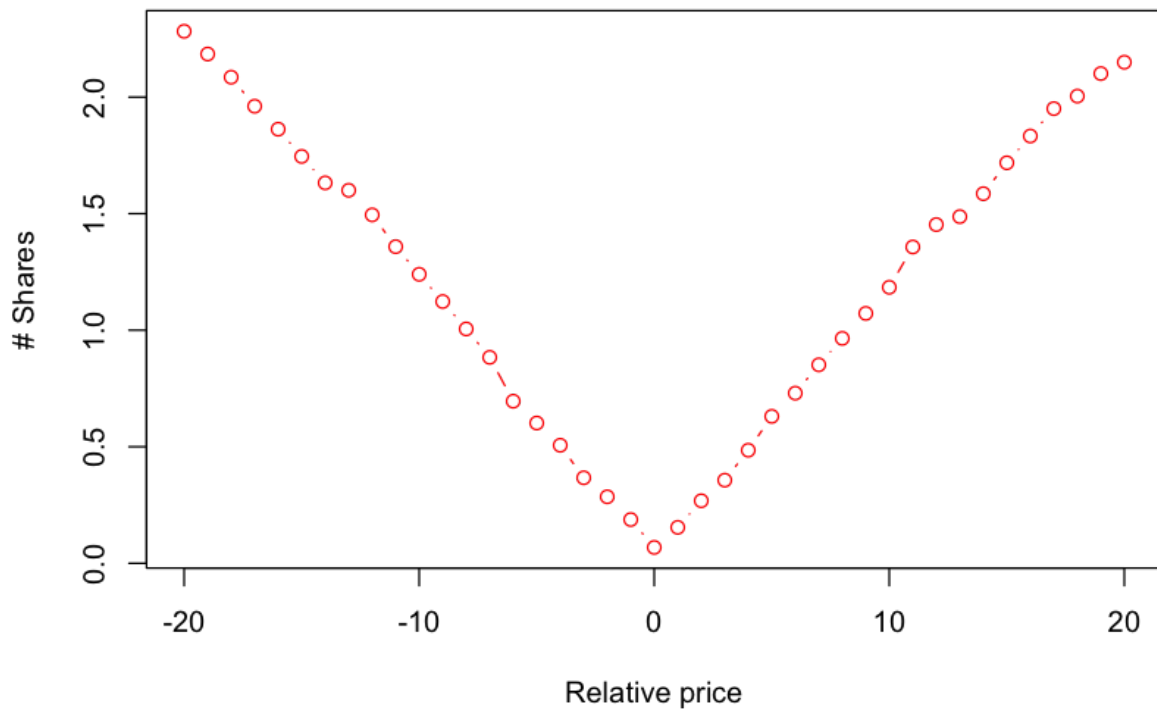


Figure 2: Average book shape over 100,000 events with  $\lambda = 1$ ,  $\mu = 10$ ,  $\nu = 1/5$ . Note that asymptotic book depth is  $\lambda/\nu = 5$ .

### Price signal in the ZI simulation

- Even in the ZI model, the shape of the order book allows prediction of price movements.
  - Traders really would need to have zero intelligence not to condition on book shape!



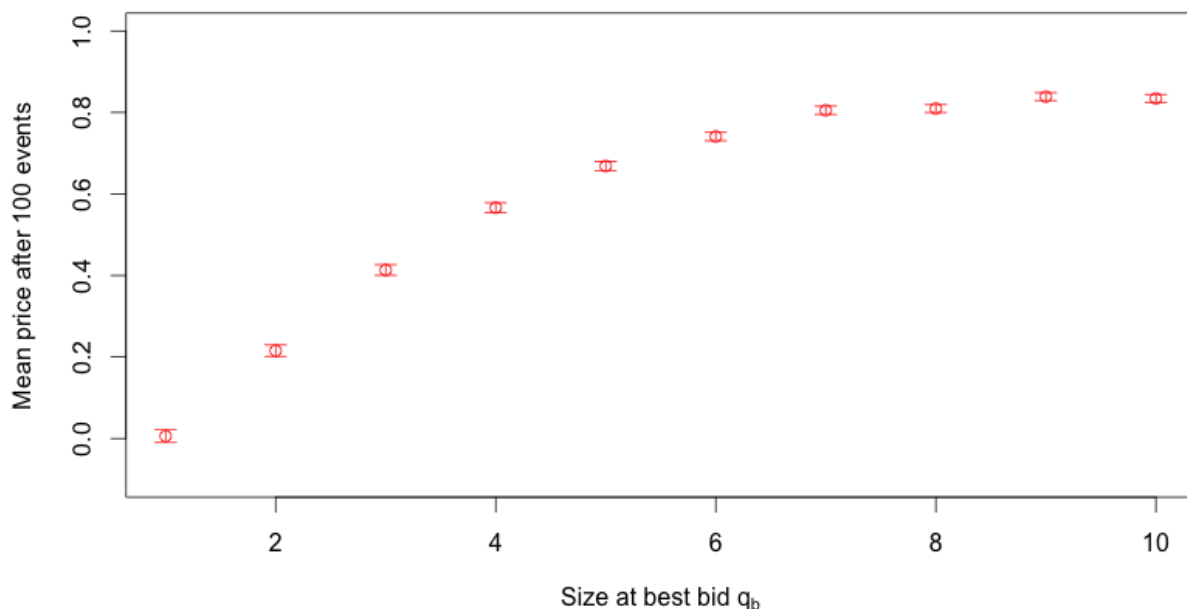


Figure 3: With one share at best offer, future price change vs size at best bid.

### Mean-reversion in the Santa Fe model

- As explained in Section 8.5.3 of [Trades, Quotes and Prices]<sup>[2]</sup>, the mid price mean-reverts in the Santa Fe model.
- To see this, suppose the price moves because the best bid queue is depleted.
  - This level in the order book will refill with either a limit buy order or a limit sell order.
  - If a limit buy, we are back in the old position.
  - If a limit sell, the price moves down.
    - However, in this case, there will always be an order imbalance consistent with a future up move.
    - Orders at the new new best bid have been shielded from market orders, and there is only one order on the ask.

### Market making in the Santa Fe model

- A related feature of the model is that market making would be highly profitable.
- However, were a market maker to be introduced, it would no longer be a zero-intelligence model.
  - Trader behavior would become strategic.
- This suggests a whole class of more sophisticated *minimal intelligence* models.

- One improvement might be to bias the probabilities of new order types depending on the order imbalance.

## Further developments in ZI models

- [Cont, Stoikov and Talreja]<sup>[4]</sup> analyze a generalized Santa Fe (ZI) model as a Markov chain.
  - They show how to compute conditional probabilities using Laplace transform techniques. For example:
    - The probability that the mid-price increases at its next move.
      - Depends on the order book imbalance.
    - The probability of executing an order before the mid-price moves.
    - The probability of making the spread.
      - Of interest to market makers. Cancel limit orders if price is going to move!
- [Cont, Stoikov and Talreja]<sup>[5]</sup> choose a limit order arrival ("rain") distribution and a cancellation distribution (by  $\Delta$ ) so as to match observed average book shapes for a Japanese stock.

## Cont and de Larrard

[Cont and de Larrard]<sup>[2]</sup> consider the following caricature of the ZI model, appropriate for large tick stocks:

- Only quantities at the best bid and the best ask ( $q_a$  and  $q_b$  respectively) are considered.
- All order sizes are for one unit. Cancellations are not proportional to existing quantity.
- $q_a$  and  $q_b$  are independent.
- The price moves down if  $q_b \downarrow 0$  and up if  $q_a \downarrow 0$ .
- After a price change,  $q_a$  and  $q_b$  are picked from a stationary distribution of queue sizes, independent of history (which could for example be empirically determined).

Formally,

$$q_a(t) = N_t^{LS} - N_t^{MB} - N_t^{CA}$$

$$q_b(t) = N_t^{LB} - N_t^{MS} - N_t^{CB}$$

where

- $N_t^{LS}$  and  $N_t^{LB}$  are Poisson processes with intensity  $\lambda$  corresponding to limit orders
- $N_t^{MS}$  and  $N_t^{MB}$  are Poisson processes with intensity  $\mu$  corresponding to market orders

- $N_t^{CA}$  and  $N_t^{CB}$  are Poisson processes with intensity  $\delta$  corresponding to cancelation of limit orders

All the processes are independent.

Hence,  $q_a(t)$  and  $q_b(t)$  are compound Poisson processes with intensity  $\Lambda := \lambda + \mu + \delta$  and jump size  $J$  distributed as

$$J = \begin{cases} 1 & \text{with probability } \frac{\lambda}{\lambda + \mu + \delta}; \\ -1 & \text{with probability } \frac{\mu + \delta}{\lambda + \mu + \delta}. \end{cases}$$

In a second paper, [Cont and de Larrard]<sup>[3]</sup> derive closed-form asymptotic expressions for the following quantities:

- Distribution of duration until the next price move.
- Probability of upward move in the price.
- Autocorrelation of price changes.
- Price volatility.

## Probability of price move up

Cont and de Larrard derive the following asymptotic expression assuming a balanced book (a book where the arrival rate of limit orders balances the arrival rates of market orders and cancelations):

### Proposition (Proposition 2 of Cont & de Larrard)

The probability  $\phi(n, p)$  that the next price is an increase, conditioned on having  $n$  orders at the bid and  $p$  orders at the ask is:

$$\phi(n, p) = \frac{1}{\pi} \int_0^\pi \left( 2 - \cos t - \sqrt{(2 - \cos t)^2 - 1} \right)^p \frac{\sin n t \cos \frac{t}{2}}{\sin \frac{t}{2}} dt$$

- This leads to a natural expression for microprice in large tick markets!

## Infinitesimal generator

The infinitesimal generator of the (two dimensional) process  $q_a(t)$  and  $q_b(t)$  is

$$\begin{aligned} \mathcal{L}u(x, y) = & \lambda[u(x + h, y) + u(x, y + h) - 2u(x, y)] \\ & + (\mu + \nu)[u(x - h, y) + u(x, y - h) - 2u(x, y)] \end{aligned}$$

where  $h$  is the tick size.

On average, we must have a balanced order book so that  $\lambda = \mu + \nu$ . Then  $\lambda = \mu + \nu = \frac{\Lambda}{2}$ .  $\mathcal{L}$  becomes

$$\mathcal{L}u(x, y) = \frac{\Lambda}{2} \left[ u(x + h, y) + u(x, y + h) + u(x - h, y) + u(x, y - h) - 4u(x, y) \right].$$

## Boundary value problem for probability of up move

Denote  $\tau_a = \inf\{t \geq 0 : q_a(t) = 0\}$  and  $\tau_b = \inf\{t \geq 0 : q_b(t) = 0\}$ . In other words,  $\tau_a$  ( $\tau_b$  respectively) is the first time that the ask (bid respectively) side of the book is depleted.

Let  $u(x, y) = \mathbb{P}[\tau_a \leq \tau_b | q_b = x, q_a = y]$ . That is,  $u$  is the probability that, conditioned on  $q_b = x$  and  $q_a = y$ , the ask side is depleted before the bid side. Consequently,  $u$  is the probability that the price moves up.

Then  $u$  satisfies the partial difference equation

$$\begin{aligned} \mathcal{L}u(x, y) &= u(x + h, y) + u(x, y + h) + u(x - h, y) + u(x, y - h) - 4u(x, y) \\ &= 0, \quad \forall x, y > 0 \end{aligned}$$

with boundary conditions

$$u(x, 0) = 1, \quad \forall x \geq 0 \quad \text{and} \quad u(0, y) = 0, \quad \forall y \geq 0.$$

## A heuristic asymptotic analysis

Recall the difference equation for  $u$ :

$$u(x + h, y) + u(x, y + h) + u(x - h, y) + u(x, y - h) - 4u(x, y) = 0.$$

Rewrite the last equation as

$$\begin{aligned} &\frac{1}{h^2} [u(x + h, y) + u(x - h, y) - 2u(x, y)] \\ &+ \frac{1}{h^2} [u(x, y + h) + u(x, y - h) - 2u(x, y)] = 0. \end{aligned}$$

In the limit, as  $h \rightarrow 0$ , the difference equation converges to the Laplace equation

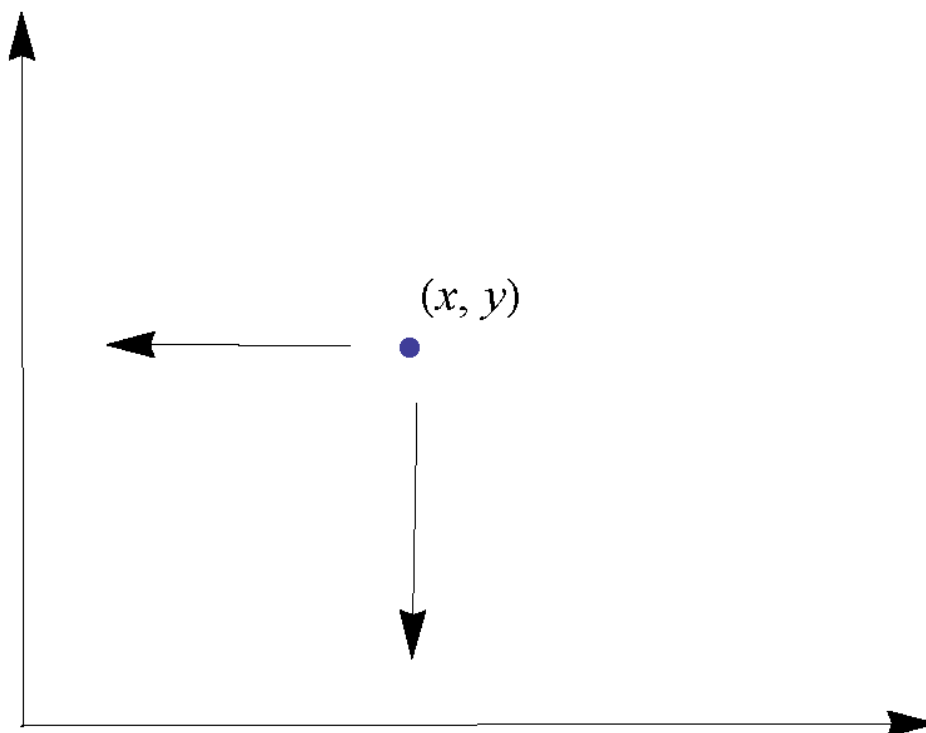
(1)

$$\Delta u = u_{xx} + u_{yy} = 0.$$

- For small  $h$ , we would expect the solution of this PDE to be approximately equal to the solution of the difference equation.

## Probability that next move is up

Let  $\phi(x, y)$  be the probability that the next move is up given a bid size of  $x$  and an ask size of  $y$ . Then  $\phi$  satisfies the Laplace equation (1) with boundary conditions with boundary condition  $\phi(x, 0) = 1$  and  $\phi(0, y) = 0$ .



### Solving the Laplace equation for $\phi$

Rewriting (1) in polar coordinates  $(r, \theta)$  gives

$$\Delta\phi(x, y) = \frac{\partial^2\phi}{\partial r^2} + \frac{1}{r} \frac{\partial\phi}{\partial r} + \frac{1}{r^2} \frac{\partial^2\phi}{\partial\theta^2}.$$

Radial symmetry gives  $\partial_r\phi = 0$  so this equation simplifies to

$$\frac{\partial^2\phi}{\partial\theta^2} = 0.$$

with solution  $\phi(r, \theta) = a\theta + b$ .

Applying the boundary conditions,  $\phi(r, \pi/2) = 0$  and  $\phi(r, 0) = 1$  gives

$$\phi(x, y) = \phi(r, \theta) = 1 - \frac{2}{\pi} \theta = 1 - \frac{2}{\pi} \arctan \frac{y}{x} = \frac{2}{\pi} \arctan \frac{x}{y}.$$

## Asymptotic vs Cont and de Larrard formula

We set  $y = 1$  and plot  $\phi(x, 1)$  vs  $x$ .

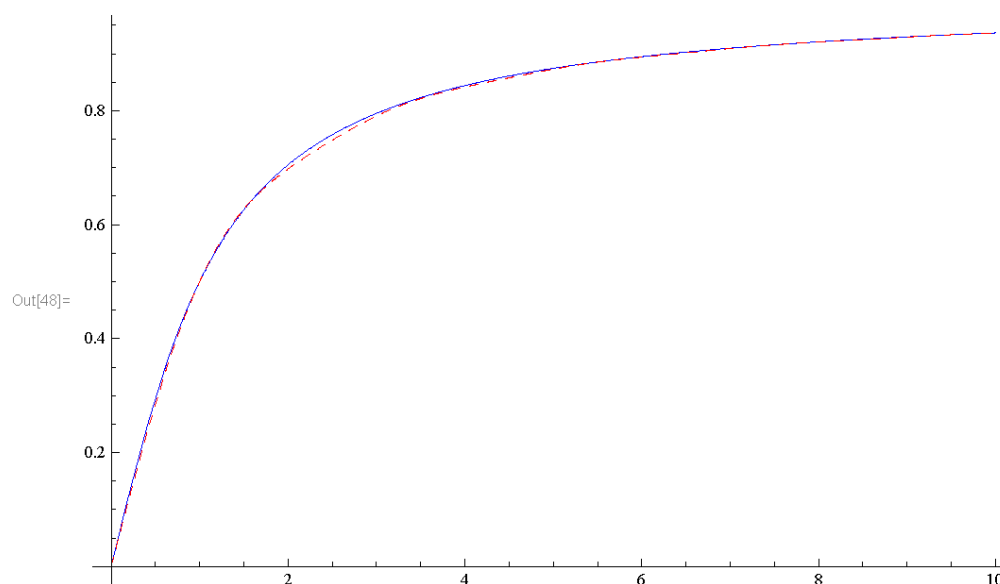


Figure 4: The blue line is simply  $\frac{2}{\pi} \arctan \frac{x}{y}$ ; the red dashed line is the Cont and de Larrard formula  $\frac{1}{\pi} \int_0^\pi dt \left( 2 - \cos t - \sqrt{(2 - \cos t)^2 - 1} \right)^y \frac{\sin x t \cos \frac{t}{2}}{\sin \frac{t}{2}}$

## Microprice

Now that we have the probability of an up-tick, we may define a quantity which we may consider as an estimate of the efficient price.

Denote the spread by  $s$  and the mid-price by  $m_t = (b_t + a_t)/2$ . Then we may define the *microprice*  $\tilde{s}_t$  as

(2)

$$\begin{aligned} \tilde{s}_t &= \text{mid-price} + \text{half spread} \times \text{expected price move} \\ &= m_t + \frac{s}{2} \{ 2 \phi(q_b, q_a) - 1 \} \\ &\approx M_t + \frac{s}{2} \left\{ \frac{4}{\pi} \arctan \frac{q_b}{q_a} - 1 \right\}. \end{aligned}$$

- If  $q_b \ll q_a$ ,  $\tilde{s}_t \approx m_t - s/2 = b_t$ .
- If  $q_b \gg q_a$ ,  $\tilde{s}_t \approx m_t + s/2 = a_t$ .

## Order book imbalance

[Gould and Bonart]<sup>[5]</sup> define the *order book imbalance*

$$I = \frac{q_b - q_a}{q_b + q_a}.$$

In terms of  $I$ , the microprice (2) becomes

$$\begin{aligned}\tilde{s}_t &\approx m_t + \frac{s}{2} \left\{ \frac{4}{\pi} \arctan \frac{q_b}{q_a} - 1 \right\} \\ &= m_t + \frac{s}{2} \frac{4}{\pi} \arctan \frac{q_b - q_a}{q_b + q_a} \\ &= m_t + \frac{s}{2} \frac{4}{\pi} \arctan I.\end{aligned}$$

- We see that book imbalance seems to be the appropriate choice of variable in the sense that  $\tilde{s}_t$  is linear in  $I$  for small  $I$ .
  - Moreover,  $I$  is anti-symmetric in  $q_b$  and  $q_a$ .

## Order book imbalance empirically

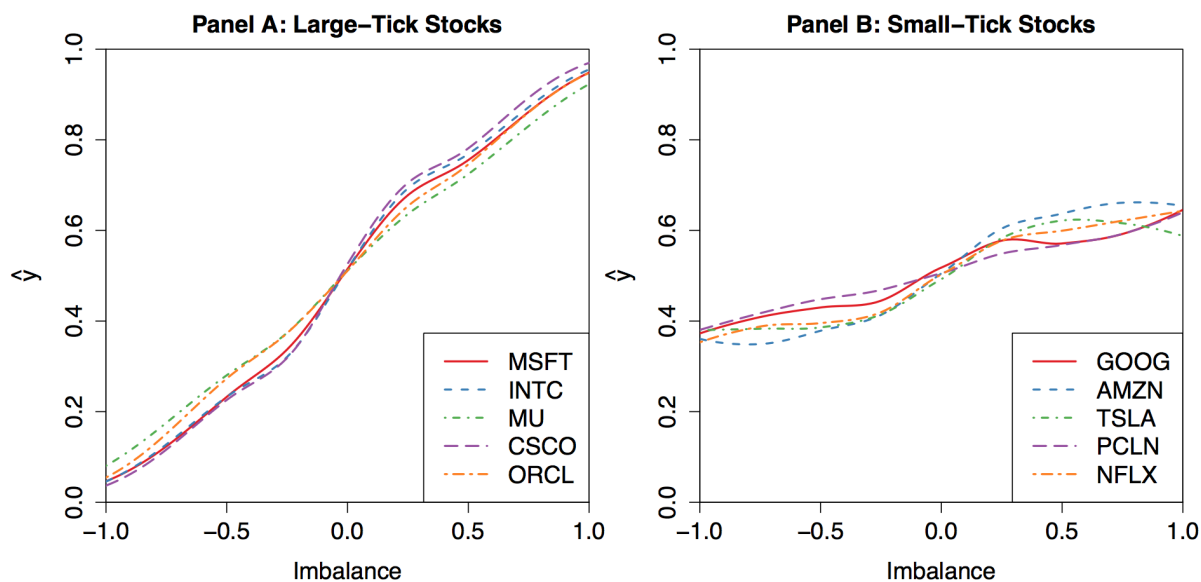


Figure 5: Empirical probability that the next mid-price moves is up from [Gould and Bonart]<sup>[6]</sup>

## Graph of microprice vs order book imbalance

```
In [22]: # Figure 6: Microprice vs order book imbalance
mPrice <- function(imbalance){4/pi*atan(imbalance)}
curve(mPrice(x),from=-1,to=1,col="red",xlab="Imbalance (I)",ylab="Microprice/Half-spread")
abline(h=0,lty=2)
abline(v=0,lty=2)
```

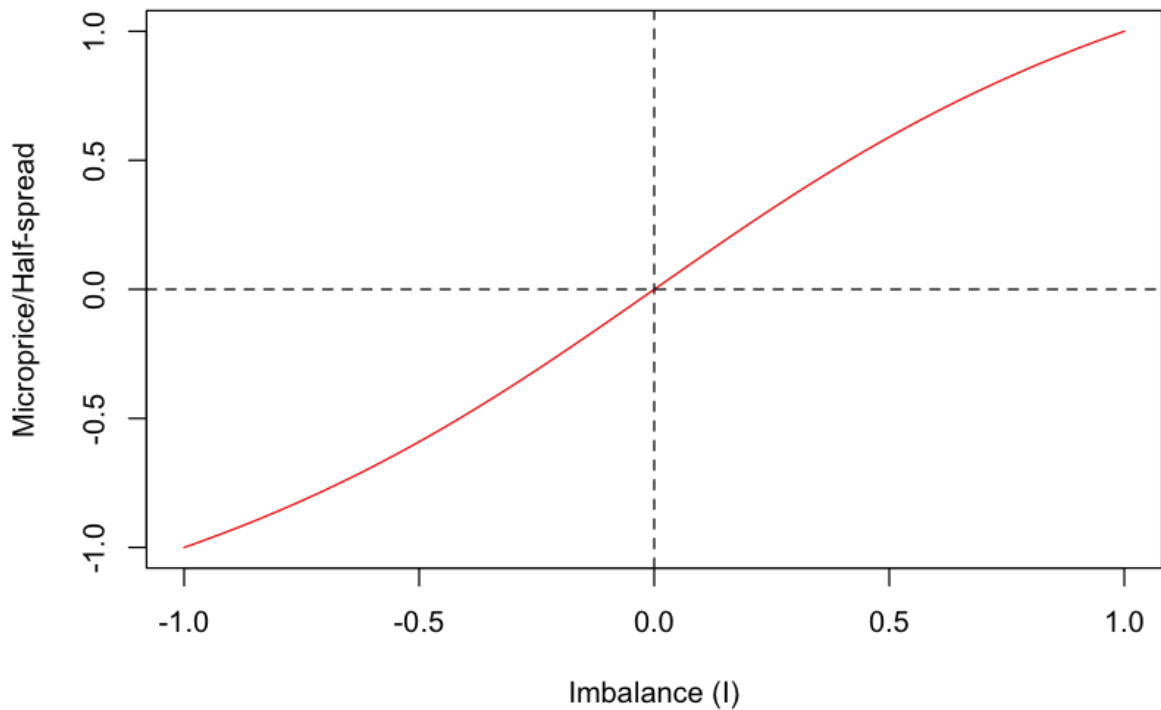


Figure 6: Microprice as a function of the order book imbalance I

- It looks nearly linear!

## Price signal in the ZI simulation vs book imbalance



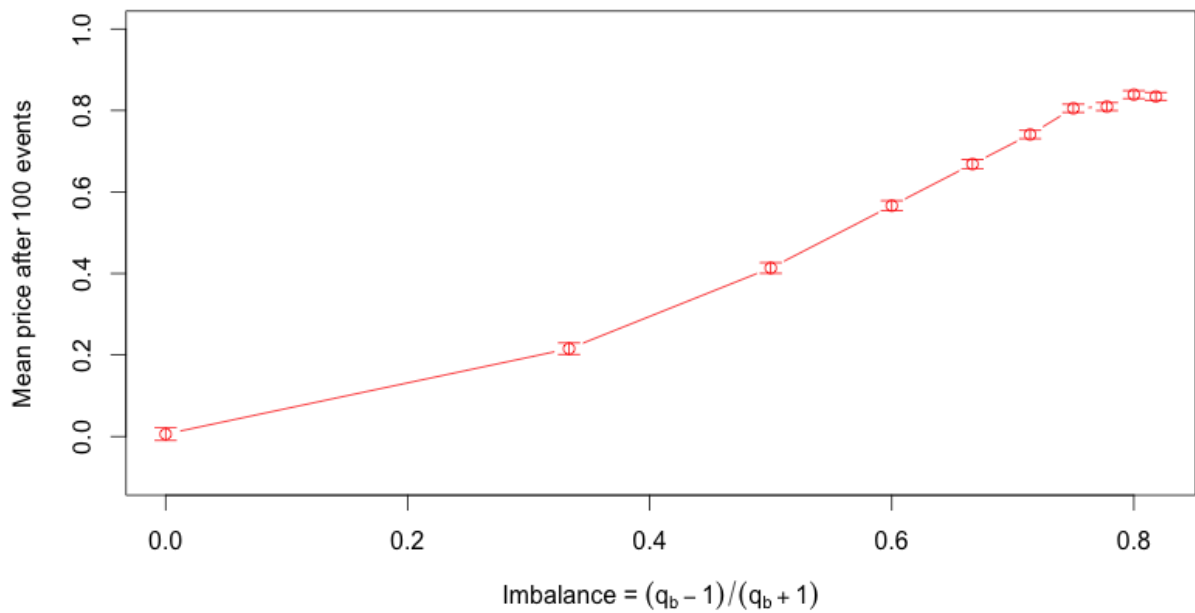


Figure 7: With one share at best offer, future price change vs book imbalance

### Price signal in the ZI simulation vs microprice

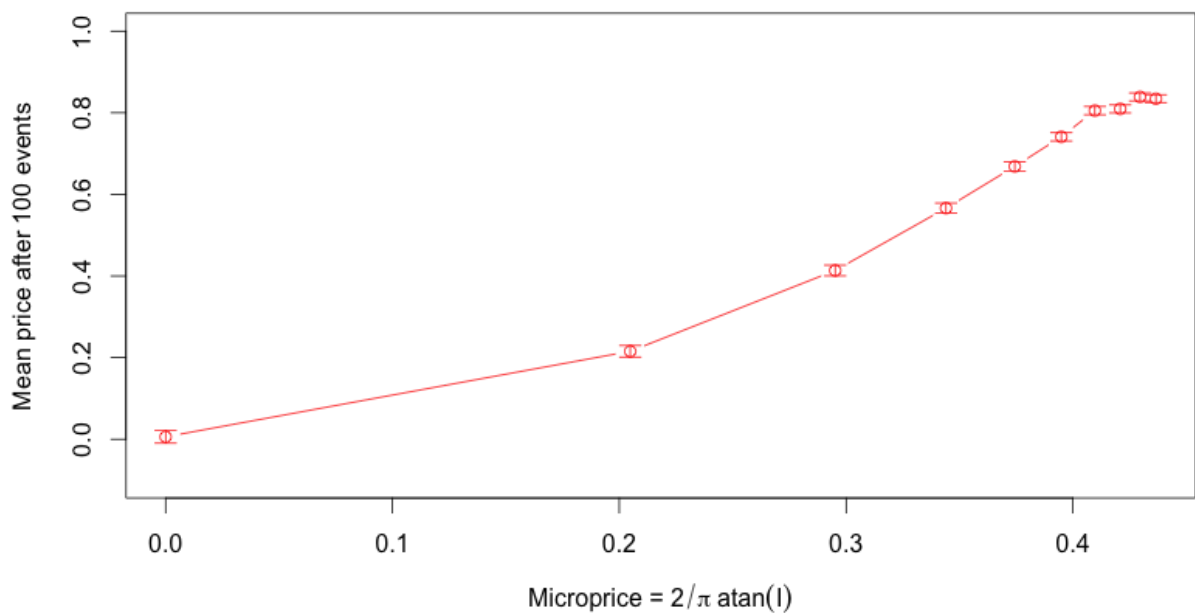


Figure 8: With one share at best offer, future price change vs microprice

## Avellaneda, Reed and Stoikov

[Avellaneda, Reed and Stoikov]<sup>[1]</sup> make the following observations about size at the best quote in real markets:

- In the zero intelligence (ZI) model, if  $q_A$  gets very small relative to  $q_B$  the probability that the next price move is up should approach 100%.
  - This is not consistent with empirical observation.
  - Avellaneda, Reed and Stoikov (ARS) ascribe this to *hidden liquidity* (this could consist of hidden orders, liquidity on other exchanges via Reg. NMS and so on).
- In the ZI model,  $q_A$  and  $q_B$  are independent.
  - In real markets  $q_A$  and  $q_B$  are negatively correlated. Indeed a good approximation is  $\rho_{AB} \approx -1$ .
  - They model this by introducing diagonal transitions which cancel an order on one side of the book and add a limit order on the other.

## The Avellaneda, Reed and Stoikov solution

Following a similar argument to the derivation of equation (1), we obtain that the probability of an up price move is given by

$$\phi(x, y; H) = u(x + H, y + H)$$

where  $H$  represents hidden liquidity and  $u$  satisfies

(2)

$$u_{xx} + \rho u_{xy} + u_{yy} = 0.$$

$H$  is interpreted as a measure of the information content of the visible order book. The smaller  $H$  is, the greater the predictive power of size at the best quote.

The solution is

$$u(x, y) = \frac{1}{2} \left\{ 1 + \frac{\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}} \frac{x-y}{x+y}\right)}{\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\right)} \right\}.$$

We then have the following special cases:

$\rho = 0$ :

$$u(x, y) = \frac{1}{2} \left\{ 1 + \frac{\arctan\left(\frac{x-y}{x+y}\right)}{\arctan 1} \right\} = \frac{1}{2} \left\{ 1 + \frac{4}{\pi} \arctan\left(\frac{x-y}{x+y}\right) \right\}$$

which is consistent with our earlier result.

and in the case  $\rho = -1$ , we have the very simple result

$$u(x, y) = \frac{x}{x + y}$$

in which case the probability of an up-move is given by

$$\phi(x, y; H) = u(x + H, y + H) = \frac{x + H}{x + y + 2H}.$$

## Empirical test

- ARS demonstrate that for large tick stocks, this simple model with  $\rho = -1$  successfully reproduces the empirical probability of an up-move.
- $H$  is typically around 15% of visible liquidity.
- By comparing  $H$  for a given stock for various exchanges, we can determine which exchange has the most informative visible order book.

## Empirical estimates of $H$

Table 4  
Implied hidden liquidity across tickers and exchanges

Ticker	NASDAQ	NYSE	BATS
XLF	0.15	0.17	0.17
QQQQ	0.21	0.04	0.18
JPM	0.17	0.17	0.11
AAPL $s = 1$	0.16	0.90	0.65
AAPL $s = 2$	0.31	0.60	0.64
AAPL $s = 3$	0.31	0.69	0.63

Table 1: Hidden liquidity  $H$  for four tickers and three exchanges.  $s$  stands for spread in ticks.

## Muni-Toke and Yoshida

A later paper by [Muni-Toke and Yoshida]<sup>[Z]</sup> updates the zero-intelligence recipe by making intensities dependent on the state of the order book, consistent with empirical estimates.

- The arrival rate of market orders depends on the spread and the size available at the opposite best quote.
- The arrival rate of limit orders depends on the total quantity (up to level 10) available on the same side of the order book.
  - Placement of the limit order is determined in a second step according to an empirically estimated distribution.

- The arrival rate of cancelations depends again on the size available and then on the position of the order in the queue.
  - Orders towards the front of the queue are less likely to be canceled.

## Order placement analysis

As discussed in [Huang, Lehalle and Rosenbaum]<sup>[6]</sup>, a realistic order book simulation can be used to estimate *ex-ante* the performance of order placement strategies. For example, one could compare

- When the mid-price changes, cancel all limit orders and place a market order.
- When the best-offer prices changes and we have the only sell order, cancel and join the queue at the next higher price level.

## Summary

- The limit order book can be viewed as a complex queuing system.
- Even with very simple rules, complex order flow and price dynamics can be generated.
- With more realistic rules, zero-intelligence models of the order book can serve as useful tools for comparing the performance of proposed order execution strategies.

## References

1. <sup>△</sup> Marco Avellaneda, Josh Reed and Sasha Stoikov, Forecasting from level-I quotes in the presence of hidden liquidity, *Alogrithmic Finance*, **1**(1) 35-43 (2011).
2. <sup>△</sup> Jean-Philippe Bouchaud, Juius Bonart, Jonathan Donier, and Martin Gould, *Trades Quotes and Prices*, Cambridge University Press (2018).
3. <sup>△</sup> Rama Cont and Adrien de Larrard, Price dynamics in a Markovian limit order market, *SIAM J. Finan. Math.*, **4**(1) 1-25 (2013).
4. <sup>△</sup> Rama Cont and Adrien de Larrard, Order book dynamics in liquid markets: limit theorems and diffusion approximations, *SSRN*, (2011).
5. <sup>△</sup> Rama Cont, Sasha Stoikov and Rishi Talreja, A stochastic model for order book dynamics, *Operations Research*, **58** 549-563 (2010).
6. <sup>△</sup> Martin Gould and Julius Bonart, Queue imbalance as a one-tick-ahead price predictor in a limit order book, *Market Microstructure and Liquidity* **2**(2) 1650006 (2016).
7. <sup>△</sup> Weibing Huang, Charles-Albert Lehalle and Mathieu Rosenbaum, Simulating and analyzing order book data: The queue-reactive model, *Journal of the American Statistical Association*

**110**(509) 107-122 (2015).

8. ^ loane Muni Toke and Nakahiro Yoshida, Modelling intensities of order flows in a limit order book, *Quantitative Finance* **17**(5) 683-701 (2017).

In [ ]: