MTH 9879 Market Microstructure Models: Spring 2024

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

Lecture note by Jim Gatheral

Outline of Lecture 1

- Overview of the course

- Market structure

    - Order types

    - Market types

    - Trading venues

- 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.
- *Order-driven* markets
  - 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/.

BATS (BZX) order book for MSFT

MSFT 🔍			Orders Accepted	Total Volume
MICROSOFT CORP COM			805,794	1,554,142

	TOP OF BOOK			LAST 10 TRADES	
	Shares	Price	Time	Price	Shares
ASKS	602	163.03	13:30:12	162.98	200
	500	163.02	13:30:09	163.00	95
	500	163.01	13:30:09	163.00	105
	500	163.00	13:30:09	163.00	495
	400	162.99	13:30:09	163.00	50
BIDS	200	162.97	13:30:09	163.00	50
	400	162.96	13:30:09	163.00	3
	300	162.95	13:30:09	163.00	2
	300	162.94	13:30:07	162.99	150
	302	162.93	13:30:06	162.99	50

Last updated 13:30:16

BATS (BYX) order book for MSFT

MSFT 🔍			Orders Accepted	Total Volume
MICROSOFT CORP COM			460,976	269,868

	TOP OF BOOK			LAST 10 TRADES	
	Shares	Price	Time	Price	Shares
ASKS	200	164.00	13:30:43	163.02	10
	100	163.50	13:29:55	162.99	100
	28	163.43	13:29:47	163.00	100
	100	163.13	13:29:46	163.01	100
	20	163.12	13:29:22	163.03	100
BIDS	104	163.02	13:29:20	163.04	100
	4	163.01	13:29:18	163.05	100
	4	163.00	13:29:08	163.06	20
	16	162.92	13:29:08	163.06	80
	100	162.91	13:29:02	163.05	88

BATS (BZX) order book for GOOG

GOOG 🔍			Orders Accepted	Total Volume
ALPHABET INC CAP STK CL C			744,904	89,441

	TOP OF BOOK			LAST 10 TRADES	
	Shares	Price	Time	Price	Shares
ASKS	10	1436.14	13:30:47	1434.80	1
	5	1435.96	13:30:33	1434.94	69
	9	1435.87	13:29:44	1435.41	20
	9	1435.60	13:29:44	1435.41	100
	9	1435.33	13:27:26	1435.82	20
BIDS	75	1434.47	13:27:26	1435.82	9
	10	1434.43	13:27:26	1435.82	10
	200	1434.38	13:27:26	1435.61	19
	10	1434.16	13:27:26	1435.62	22
	100	1433.96	13:27:26	1435.82	7

Last updated 13:31:41

BATS (BYX) order book for GOOG

GOOG 🔍			Orders Accepted	Total Volume
ALPHABET INC CAP STK CL C			278,168	17,916

	TOP OF BOOK			LAST 10 TRADES	
	Shares	Price	Time	Price	Shares
ASKS	100	1506.75	13:31:16	1434.44	1
	100	1505.86	13:27:42	1435.65	5
	100	1502.70	13:22:17	1435.22	100
	100	1489.26	13:21:06	1435.10	100
	100	1435.76	13:21:05	1435.06	100
BIDS	5	1410.00	13:20:45	1434.60	100
	100	1373.76	13:19:39	1435.01	10
	100	1371.14	13:17:34	1433.95	25
	100	1363.25	13:10:44	1432.80	6
	100	1361.97	13:08:46	1433.70	100

BATS (BZX) order book for GE

GE 🔍		Orders Accepted	Total Volume
GENERAL ELECTRIC CO COM		97,965	945,921

	TOP OF BOOK			LAST 10 TRADES	
	Shares	Price	Time	Price	Shares
ASKS	4,000	11.5800	13:31:16	11.5350	13
ASKS	11,543	11.5700	13:31:00	11.5350	96
ASKS	9,533	11.5600	13:31:00	11.5300	46
ASKS	11,260	11.5500	13:31:00	11.5350	29
ASKS	3,900	**11.5400**	13:30:12	11.5350	200
BIDS	7,604	**11.5300**	13:30:05	11.5350	100
BIDS	10,200	11.5200	13:30:01	11.5350	100
BIDS	10,110	11.5100	13:29:55	11.5350	100
BIDS	11,128	11.5000	13:29:16	11.5350	100
BIDS	3,951	11.4900	13:28:52	11.5350	61

Last updated 13:32:06

BATS (BYX) order book for GE

GE 🔍		Orders Accepted	Total Volume
GENERAL ELECTRIC CO COM		57,541	697,008

	TOP OF BOOK			LAST 10 TRADES	
	Shares	Price	Time	Price	Shares
ASKS	900	11.5800	13:32:19	11.5350	100
ASKS	2,400	11.5700	13:32:19	11.5390	80
ASKS	800	11.5600	13:32:03	11.5400	27
ASKS	1,200	11.5500	13:32:03	11.5400	29
ASKS	1,075	**11.5400**	13:32:03	11.5400	100
BIDS	1,801	**11.5300**	13:31:44	11.5390	12
BIDS	2,200	11.5200	13:31:00	11.5300	151
BIDS	1,459	11.5100	13:31:00	11.5300	151
BIDS	2,368	11.5000	13:31:00	11.5300	197
BIDS	1,000	11.4900	13:31:00	11.5300	1,103

Last updated 13:32:21

BATS (BZX) order book for FCX

Book Viewer »	**BZX Equities**	BYX Equities	EDGX Equities	EDGA Equities	Market Quality Statistics ☑

Shows the top bids and asks for any symbol.

FCX 🔍			Orders Accepted	Total Volume
FREEPORT-MCMORAN INC CL B			100,886	1,309,993

	TOP OF BOOK			LAST 10 TRADES	
	Shares	Price	Time	Price	Shares
ASKS	1,700	11.1800	13:31:46	11.1300	7
	4,100	11.1700	13:31:46	11.1300	87
	3,800	11.1600	13:31:46	11.1300	13
	3,700	11.1500	13:31:46	11.1300	100
	3,120	**11.1400**	13:31:46	11.1300	100
BIDS	1,200	**11.1300**	13:31:46	11.1300	100
	3,200	11.1200	13:31:21	11.1250	200
	5,400	11.1100	13:31:21	11.1200	9
	5,700	11.1000	13:31:19	11.1200	64
	3,200	11.0900	13:31:19	11.1200	27

Last updated 13:32:56

BATS (BYX) order book for FCX

Book Viewer »	BZX Equities	**BYX Equities**	EDGX Equities	EDGA Equities	Market Quality Statistics ☑

Shows the top bids and asks for any symbol.

FCX 🔍			Orders Accepted	Total Volume
FREEPORT-MCMORAN INC CL B			56,303	846,541

	TOP OF BOOK			LAST 10 TRADES	
	Shares	Price	Time	Price	Shares
ASKS	800	11.18	13:32:36	11.13	100
	1,800	11.17	13:31:50	11.13	100
	600	11.16	13:31:46	11.13	100
	700	11.15	13:31:46	11.13	100
	1,100	11.14	13:31:46	11.13	100
BIDS	700	11.12	13:31:46	11.13	266
	1,000	11.11	13:31:46	11.13	5
	2,100	11.10	13:31:46	11.13	29
	600	11.09	13:31:46	11.13	300
	500	11.08	13:31:46	11.13	100

Last updated 13:32:41

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/ and here 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 were large tick stocks at the time this slide was written (https://www.yahoo.com/now/reverse-stock-split-ge-trading-145334436.html).
    - 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 μ.
- Limit orders (per price level) arrive at rate λ.
- A proportion ν 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 λ/ν.

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 Δ, 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,

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

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 λ/ν.
    - 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
μ	shares/T
λ	shares/(ticks T)
ν	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	μ / ν
T	$1 / \nu$
ticks	μ / λ

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 λ/ν.
- As we saw earlier, the exact result is in fact just λ/ν 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 μ/λ.
- 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 μ.
- We deduce that (roughly at least)

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

The slope of the order book density

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

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

- Thus the (dollar) 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, μ/ν is the only combination of parameters with dimensions of $shares$.

Market impact

The most familiar market impact model is $\sigma \sqrt{L}$ where σ is the (dollar) 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 μ/λ.
- 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 $f(.)$.

- 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.
- 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 μ, λ and ν 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 λ, cancelation rate is ν 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

Event	Notation
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

$$\{\mathbb{P}_{MB}, \mathbb{P}_{MS}, \mathbb{P}_{LB}, \mathbb{P}_{LS}, \mathbb{P}_{CS}, \mathbb{P}_{CB}\} \propto \{\mu/2, \mu/2, L\,\lambda, L\,\lambda, \nu\,n_A, \nu\,n_B\}.$$

− If the selected event is a limit order, draw the relative price level from $\{1, 2, \ldots, L\}$. - If the selected event is a cancelation, select randomly which order within the band to cancel.

- Update the order book and increment t.

Ouputs 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.
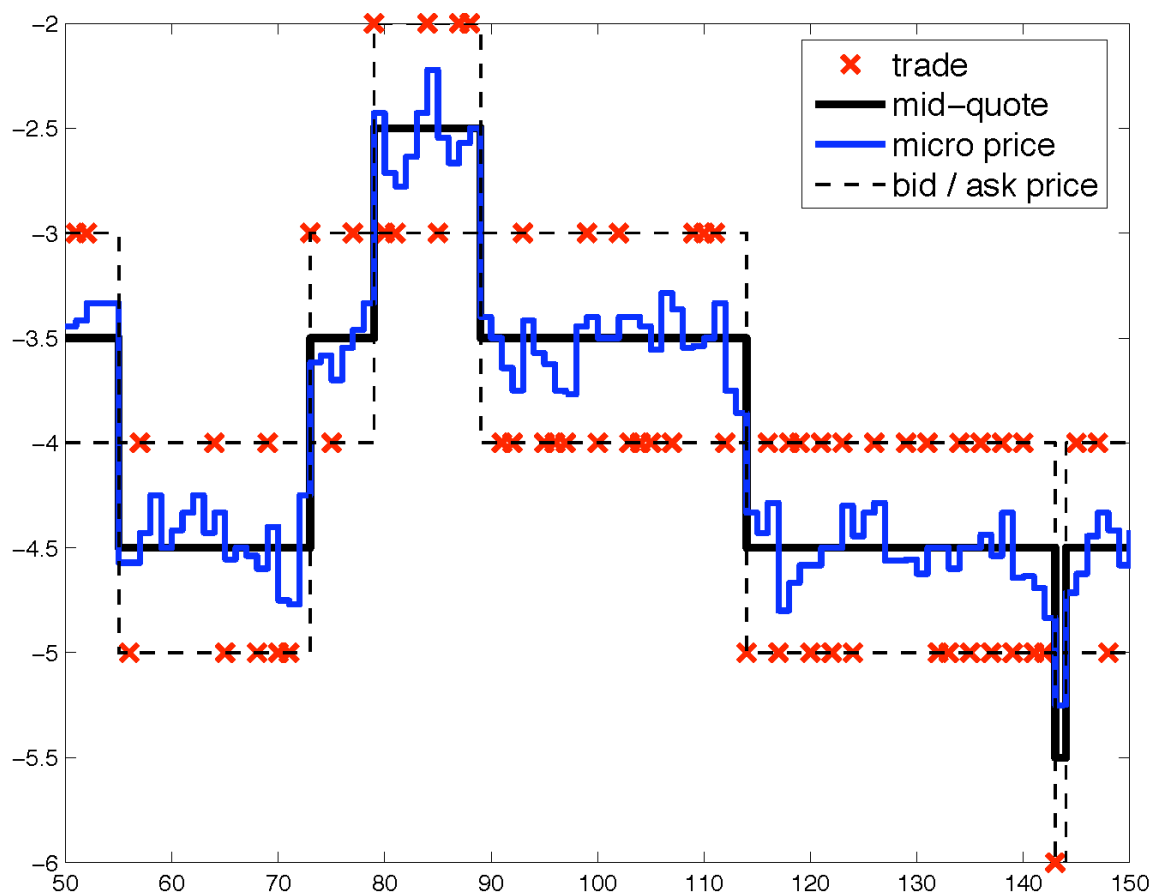
Figure 1: A typical sample path

Python implementation of ZI simulation

In [1]:
```python
import numpy as np
import random
import matplotlib.pyplot as plt
```

The zero-intelligence ZI class

In [2]:
```python
# the ZI class

class ZI:
    '''
    This defines the ZI (zero-intelligence) class
    Adapted from the R code by Jim Gatheral
    '''

    # Book setup
    L = 30 # default, number of price levels to be included in iterations
    LL = 1000 # default, total number of levels in buy and sell books
    # initial sizes on the bid/buy side
    buy_size = [5 for x in range(LL - 8)] + [5, 4, 4, 3, 3, 2, 2, 1] + [0 fo
```

```python
    buy_size = np.array(buy_size)
    # initial sizes on the ask/sell side
    sell_size = [0 for x in range(LL + 1)] + [1, 2, 2, 3, 3, 4, 4, 5] + [5 f
    sell_size = np.array(sell_size)

    def __init__(self, lmda, mu, nu, L=30, LL=1000, buy_size=buy_size, sell_
        self.lmda = lmda
        self.mu = mu
        self.nu = nu
        self.L = L
        self.LL = LL
        self.price = np.arange(-LL, LL + 1)
        self.buy_size = buy_size.copy()
        self.sell_size = sell_size.copy()

        # dict for order book
        self.book = {'price': self.price,
                     'buy_size': self.buy_size,
                     'sell_size': self.sell_size
                     }
        # market activities
        self.activities = (self.limit_buy, self.limit_sell, self.cancel_buy,
                  self.cancel_sell, self.market_buy, self.market_sell)

    def book_shape(self, band=10):
        buy = self.book['buy_size'][(self.mid_posn() - band - 1):self.mid_po
        sell = self.book['sell_size'][self.mid_posn():(self.mid_posn() + ban
        return np.concatenate([buy, sell])

    def book_shape_signed(self, band=10):
        buy = self.book['buy_size'][(self.mid_posn() - band - 1):self.mid_po
        sell = self.book['sell_size'][self.mid_posn():(self.mid_posn() + ban
        return np.concatenate([-buy, sell])

    def book_plot(self, band=10, signed=False):
        plt.figure(figsize=(8, 4))
        if signed is True:
            buy = self.book['buy_size'][(self.mid_posn() - band - 1):self.mi
            sell = self.book['sell_size'][self.mid_posn():(self.mid_posn() +
            plt.plot(np.arange(-band, 1), -buy, 'ro:')
            plt.plot([0, 1], [0, sell[0]], 'b:')
            plt.plot(np.arange(1, band + 1), sell, 'bo:')
            plt.hlines(y = 0, xmin=-band, xmax=band, linestyle='dashed', lin
            plt.title('Signed limit order book')
            plt.xlabel('Relative price')
            plt.ylabel('Quantity')
            plt.show()
            return None

        plt.plot(np.arange(-band, band + 1), self.book_shape(band), 'ro:')
        plt.xlabel('Relative price')
        plt.ylabel('Quantity')
        plt.title('Limit order book')
        plt.show()

    # functions for finding best quotes, mid, and spread
```

```python
    def best_offer(self):
        return self.book['price'][self.book['sell_size'] > 0].min()

    def best_bid(self):
        return self.book['price'][self.book['buy_size'] > 0].max()

    def spread(self):
        return self.best_offer() - self.best_bid()

    def mid(self):
        return (self.best_offer() + self.best_bid())/2

    # functions for finding positions indices
    def bid_posn(self):
        return len(self.book['buy_size'][self.book['price'] <= self.best_bid

    def ask_posn(self):
#         return len(self.book['sell_size'][self.book['price'] >= self.best_o
        return len(self.book['sell_size'][self.book['price'] <= self.best_of

    def mid_posn(self):
        return int((self.bid_posn() + self.ask_posn())/2)

    def pick(self, m):
        return random.randint(1, m)

    def go(self):
        pass

    # market activity functions
    def limit_buy(self, price=None):
        if price is None:
            prx = self.best_offer() - self.pick(self.L)
        else:
            prx = price
            if prx >= self.best_offer():
                print(f'Limit buy order at price {prx} greater than best off
                return None
        self.book['buy_size'][self.book['price']==prx] += 1
        return ['LB', prx]

    def limit_sell(self, price=None):
        if price is None:
            prx = self.best_bid() + self.pick(self.L)
        else:
            prx = price
            if prx <= self.best_bid():
                print(f'Limit sell order at price {prx} smaller than best bi
                return None
        self.book['sell_size'][self.book['price']==prx] += 1
        return ['LS', prx]

    def market_buy(self):
        prx = self.best_offer()
        self.book['sell_size'][self.book['price']==prx] -= 1
        return ['MB', prx]
```

```python
    def market_sell(self):
        prx = self.best_bid()
        self.book['buy_size'][self.book['price']==prx] -= 1
        return ['MS', prx]

    # function for finding the number of cancelable buy orders
    def n_cb(self):
        return self.book['buy_size'][self.book['price'] >= self.best_offer()

    # function for finding the number of cancelable sell orders
    def n_cs(self):
        return self.book['sell_size'][self.book['price'] <= self.best_bid()

    def cancel_buy(self, price=None):
        q = self.pick(self.n_cb())
        tmp = [x for x in self.book['buy_size']]
        tmp.reverse()
        tmp = np.array(tmp).cumsum()
        posn = len(tmp[tmp >= q]) - 1
        prx = self.book['price'][posn]
        if price is not None:
            prx = price
        self.book['buy_size'][posn] -= 1
        return ['CB', prx]

    def cancel_sell(self, price=None):
        q = self.pick(self.n_cs())
        tmp = np.cumsum(self.book['sell_size'])
        posn = len(tmp[tmp < q])
        prx = self.book['price'][posn]
        if price is not None:
            prx = price
        self.book['sell_size'][posn] -= 1
        return ['CS', prx]

    def generate_events(self, n_events=1, band=20, logging=False):
        lb_prob = self.lmda*self.L
        ls_prob = self.lmda*self.L
        mb_prob = self.mu/2
        ms_prob = self.mu/2

        # function that determines the event probabilities
        def probs():
            cb_prob = self.nu*self.n_cb()
            cs_prob = self.nu*self.n_cs()
            p = np.array([lb_prob, ls_prob, cb_prob, cs_prob, mb_prob, ms_pr
            return p/p.sum()

        avg_book_shape = np.zeros(2*band + 1)

        if logging is True:
            event_log = []
            for i in range(n_events):
                p = probs()
                act = np.random.choice(self.activities, p=p)()
```

```
                event_log.append(act)
                avg_book_shape += self.book_shape(band)/n_events

            return event_log, avg_book_shape

        for i in range(n_events):
            p = probs()
            np.random.choice(self.activities, p=p)()
            avg_book_shape += self.book_shape(band)/n_events

        return avg_book_shape
```

In [3]:
```
# parameters
# limit order arrival rate
lmda = 1

# market order arrival rate
mu = 10

# cancel order percentage arrival rate
nu = 1/5

# Instanize ZI class
zi = ZI(mu=mu, lmda=lmda, nu=nu)
```
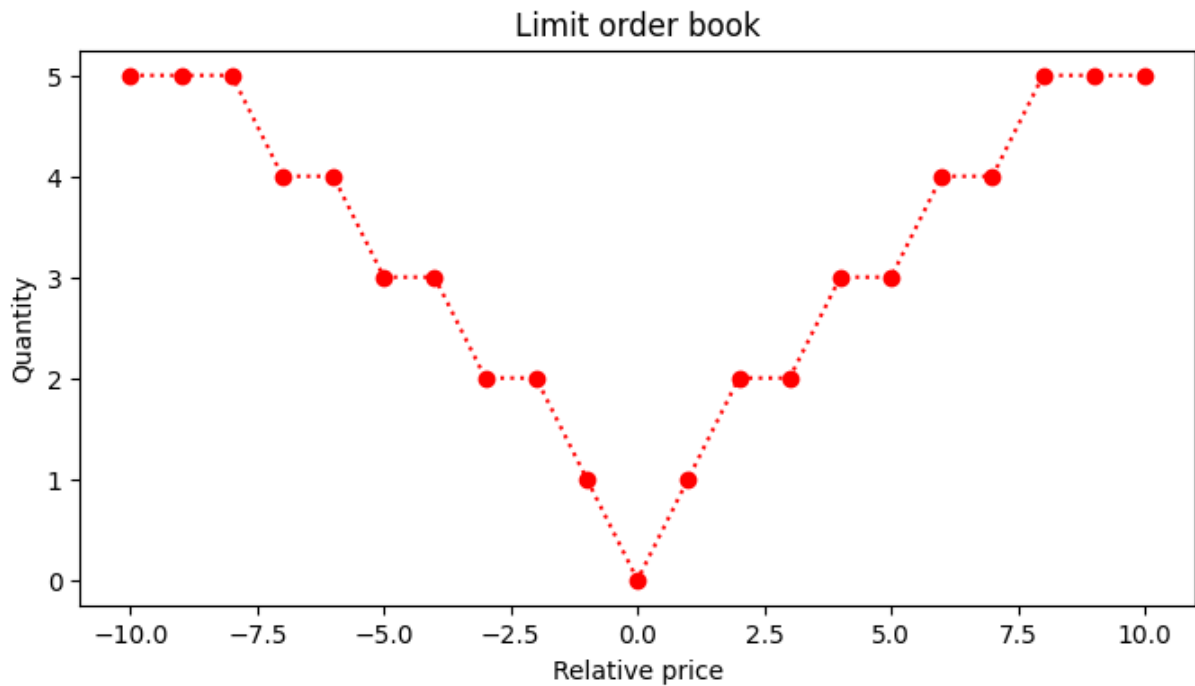
In [4]:
```
# plot order book
zi.book_plot()
zi.book_plot(signed=True)
```


Limit order book

Signed limit order book
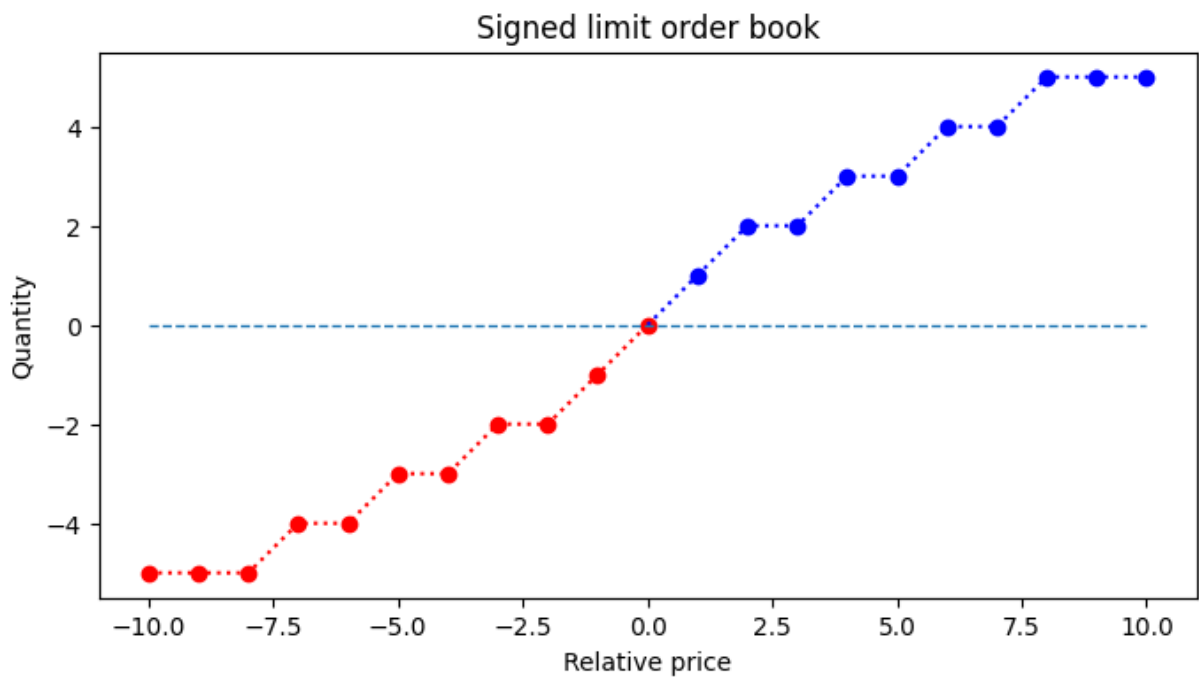
```
In [5]: zi.limit_buy(), zi.limit_sell(), zi.limit_buy()

Out[5]: (['LB', -19], ['LS', 25], ['LB', -24])

In [6]: zi.limit_buy(-2)

Out[6]: ['LB', -2]

In [7]: zi.book_plot(30)
```


Limit order book
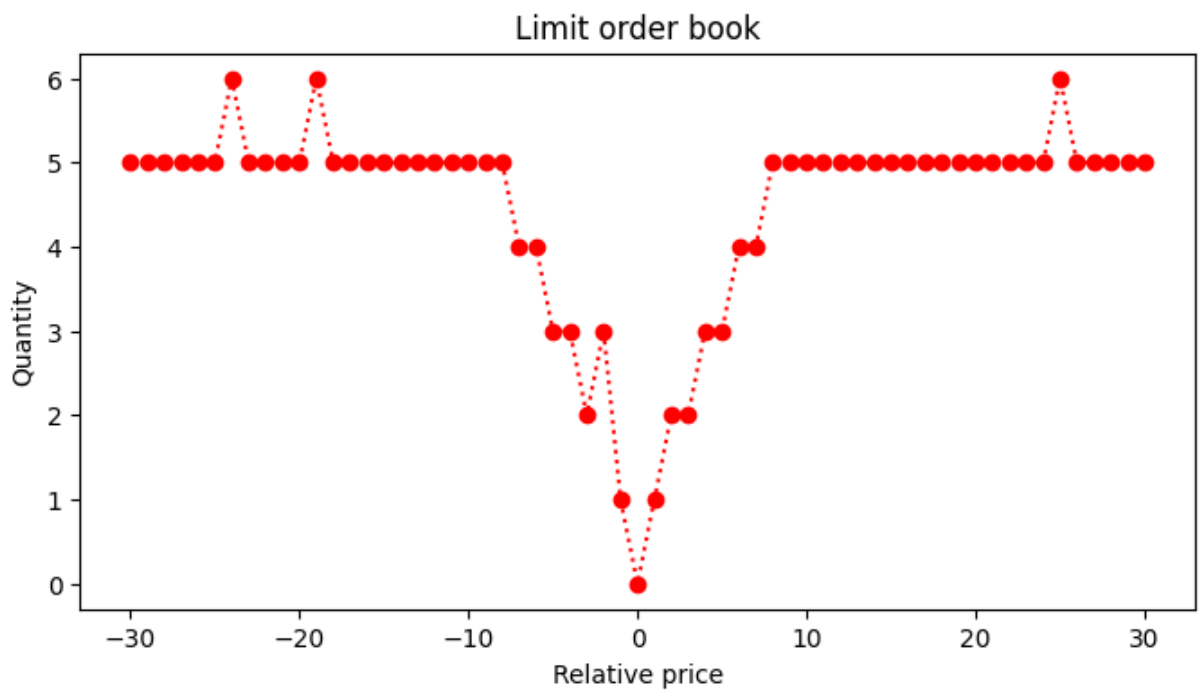
```
In [8]: # generate market activities
        zi.generate_events(logging=True)
```

```
Out[8]: ([['CS', 12]],
        array([5., 6., 5., 5., 5., 5., 5., 5., 5., 5., 5., 5., 5., 4., 4., 3., 3.,
               2., 3., 1., 0., 1., 2., 2., 3., 3., 4., 4., 5., 5., 5., 5., 4., 5.,
               5., 5., 5., 5., 5., 5., 5.]))
```

Average book shape in the ZI simulation

```python
In [9]: # Burn in for 100 events
        _ = zi.generate_events(n_events=100)

        # order book shape after 100 events
        zi.book_plot(20)
        _, zi.best_bid(), zi.mid(), zi.best_offer()
```
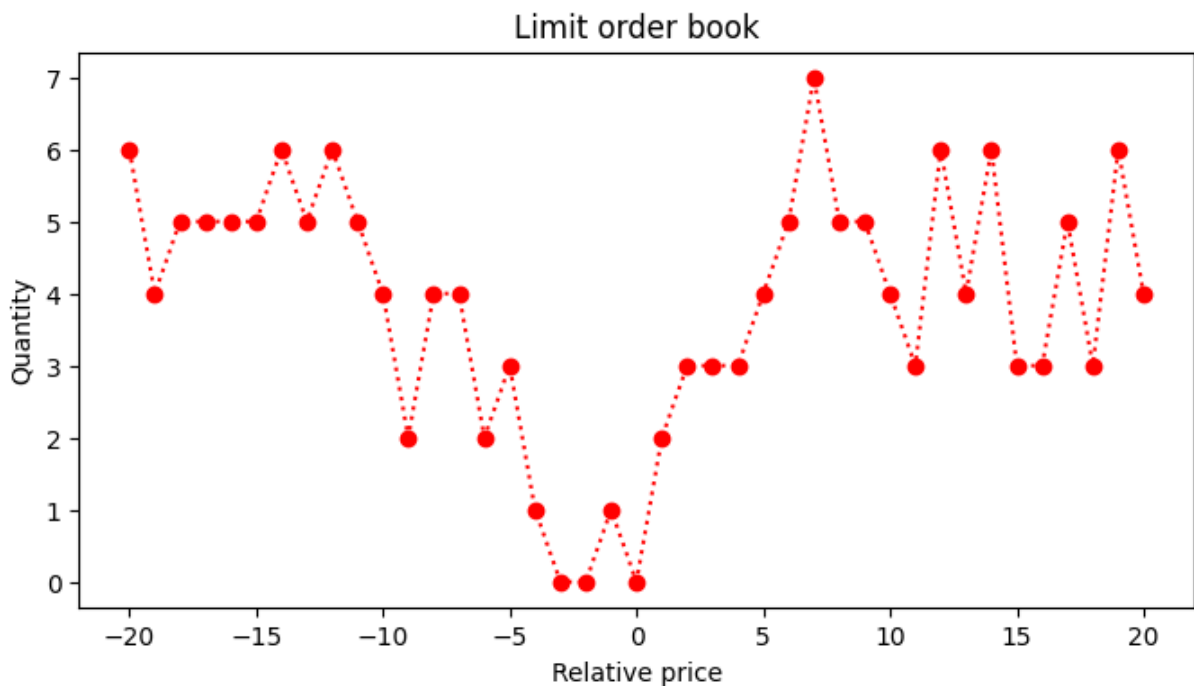


```
Out[9]: (array([5.08, 5.18, 5.04, 4.95, 4.52, 4.96, 5.15, 5.25, 5.55, 5.31, 4.93,
               4.63, 4.57, 3.95, 3.41, 3.51, 3.08, 1.94, 1.88, 0.5 , 0.2 , 0.58,
               1.83, 2.24, 2.45, 3.  , 3.43, 3.7 , 4.62, 5.08, 4.94, 4.6 , 4.15,
               4.25, 4.02, 4.57, 3.65, 4.38, 4.21, 4.31, 4.83]),
        2,
        3.0,
        4)
```

```python
In [10]: import time

         start = time.time()
         n_events = 100_000
         avg_book_shape = zi.generate_events(n_events=n_events)
         elapsed_time = time.time() - start
         print(elapsed_time)
         print(time.strftime("%H:%M:%S", time.gmtime(elapsed_time)))
         avg_book_shape, zi.mid()
```

```
39.38439130783081
00:00:39
```

(array([4.87109, 4.80249, 4.85763, 4.95798, 4.99814, 4.93154, 4.95633,
        4.86352, 4.79453, 4.74312, 4.72465, 4.63295, 4.50165, 4.28756,
        3.96353, 3.57674, 3.13647, 2.63112, 1.97641, 1.23473, 0.41787,
        0.73913, 1.34001, 1.94366, 2.45859, 3.02062, 3.51301, 4.00209,
        4.35866, 4.59263, 4.8567 , 4.96217, 4.98094, 5.05743, 4.99626,
        4.97851, 4.90948, 4.86007, 4.88755, 4.95177, 5.03056]),
  -21.0)

```python
# plot average book shape
band = 20
plt.figure(figsize=(9, 5))
plt.plot([x for x in range(-band, band + 1)], avg_book_shape, 'b:')
plt.plot([x for x in range(-band, band + 1)], avg_book_shape, 'ro')
plt.title(f'Avage book shape after {n_events} events')
plt.xlabel('Relative price')
plt.ylabel('Quantity');
```
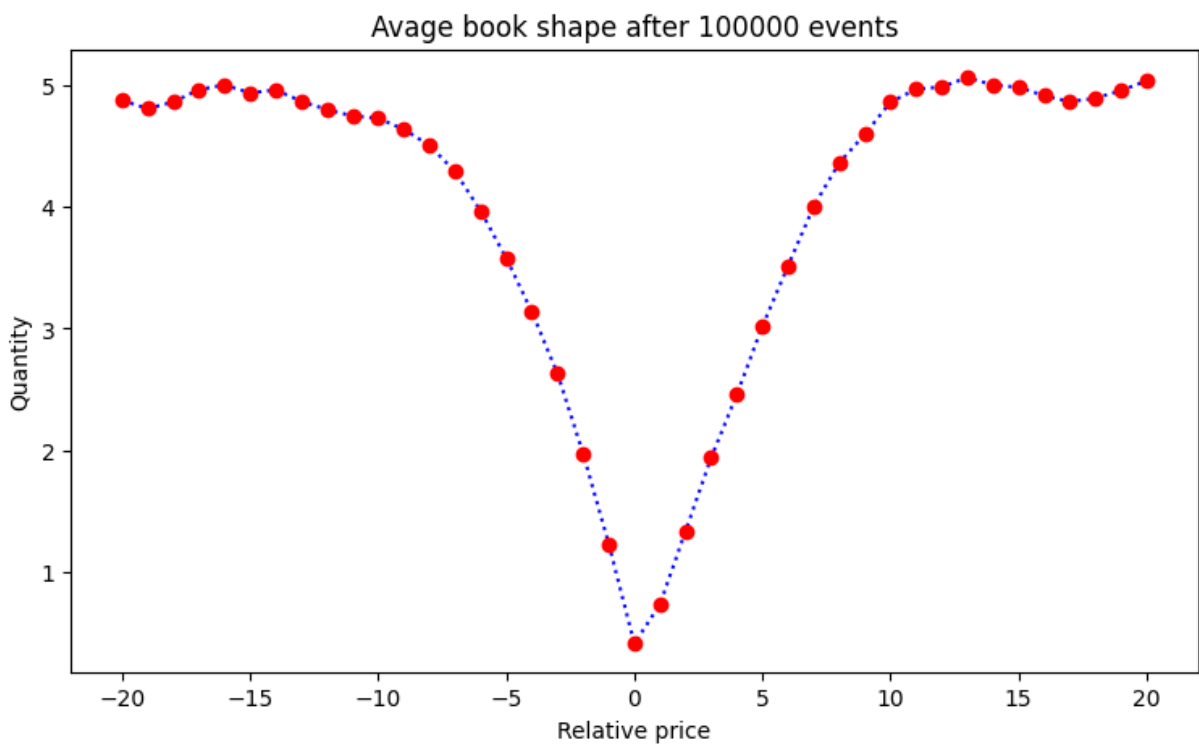

Figure 2: Average book shape over 100,000 events with $\lambda = 1$, $\mu = 10$, $\nu = 1/5$. Note that asymptotic book depth is $\lambda/\nu = 5$.

Price signal in the ZI simulation

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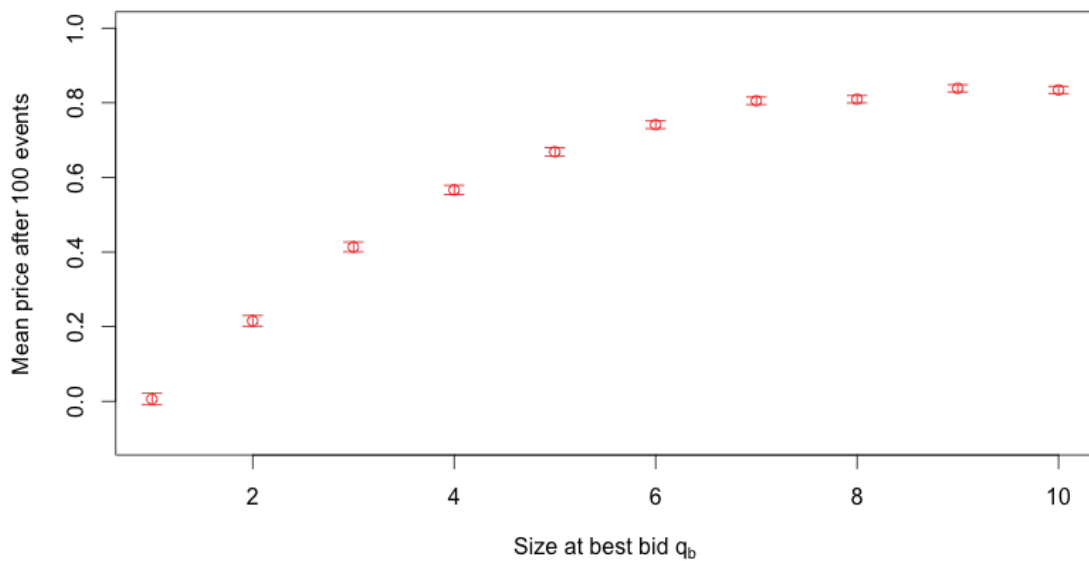 - Traders really would need to have zero intelligence not to condition on book shape!

Figure 3: With one share at best offer, average price change after 100 events vs size at best bid. Each point is generated from 10,000 simluations; error bars are computed as 2 standard deviations of the mean.

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 sophisiticated *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][5] 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][5] choose a limit order arrival ("rain") distribution and a cancelation distribution (by Δ) so as to match observed average book shapes for a Japanese stock.

Cont and de Larrard

[Cont and de Larrard][3] 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. Cancelations 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 λ corresponding to limit orders
- N_t^{MS} and N_t^{MB} are Poisson processes with intensity μ corresponding to market orders
- N_t^{CA} and N_t^{CB} are Poisson processes with intensity ν 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][4] 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

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

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

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

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, τ_a (τ_b respectively) is the first time that the ask (bid respectively) side of the book is depleted.

Let $u(x, y) = \mathbb{P}\left[\tau_a \leq \tau_b | q_b = x, q_a = y\right]$. 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

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

with boundary conditions

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

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

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

In the limit, as $h \to 0$, 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 ϕ satisfies the Laplace equation (1) with boundary conditions with boundary condition $\phi(x,0) = 1$ and $\phi(0,y) = 0$.
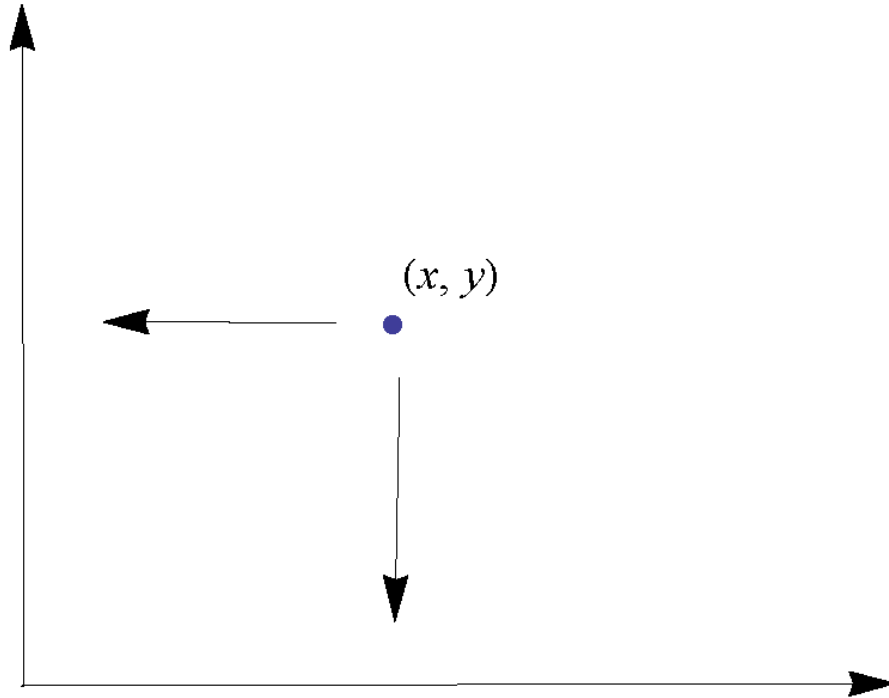
Solving the Laplace equation for ϕ

Rewriting (1) in polar coordinates (r, θ) 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
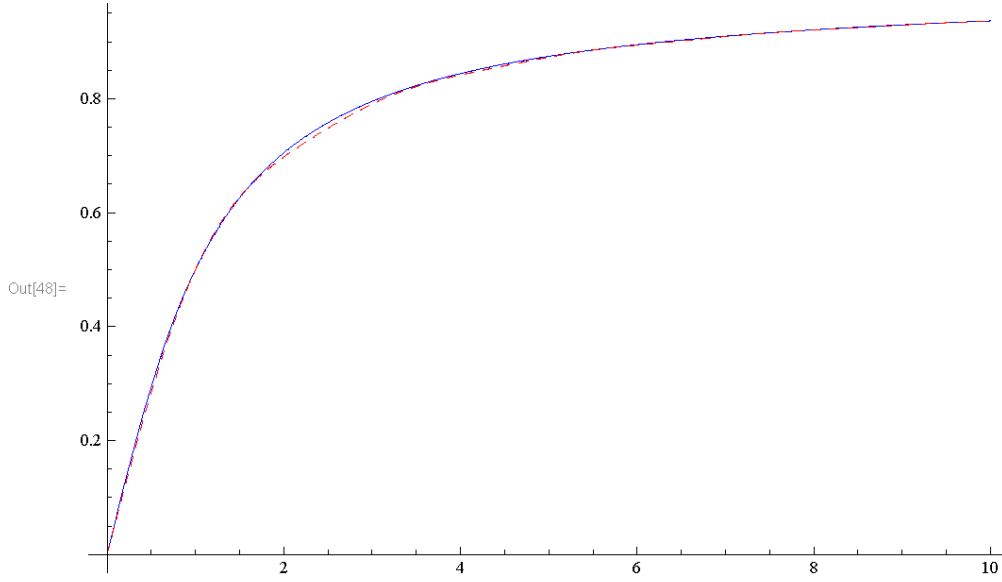
We set $y = 1$ and plot $\phi(x, 1)$ vs x.

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)

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

- 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][6] define the *order book imbalance*

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

In terms of I, the microprice (2) becomes

$$\tilde{s}_t \approx m_t + \frac{s}{2}\left\{\frac{4}{\pi}\arctan\frac{q_b}{q_a} - 1\right\} \tag{3}$$

$$= m_t + \frac{s}{2}\frac{4}{\pi}\arctan\frac{q_b - q_a}{q_b + q_a} \tag{4}$$

$$= m_t + \frac{s}{2}\frac{4}{\pi}\arctan I. \tag{5}$$

- 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.

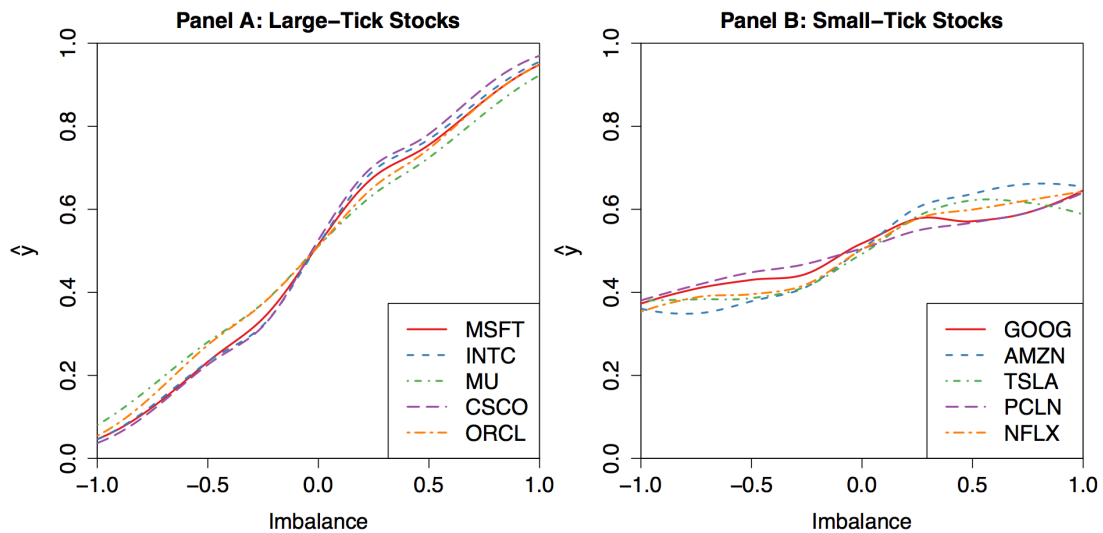## Order book imbalance empirically

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

Graph of microprice vs order book imbalance

```
In [12]:  # Figure 6: Microprice vs order book imbalance
          def micro_price(imbalance):
              price = 4/np.pi*np.arctan(imbalance)
              return price


          x = np.linspace(-1, 1, 1000)
          y = micro_price(x)

          plt.figure(figsize=(9, 5))
          plt.plot(x, y, color='red')
          plt.hlines(y=0, xmin=x.min(), xmax=x.max(), linewidth=1, linestyle='dashed')
          plt.vlines(x=0, ymin=y.min(), ymax=y.max(), linewidth=1, linestyle='dashed')
          plt.title('Micro-price vs Imbalance')
```

```
plt.xlabel('Imbalance (I)')
plt.ylabel('Micro-price/Half-spread')
plt.show()
```
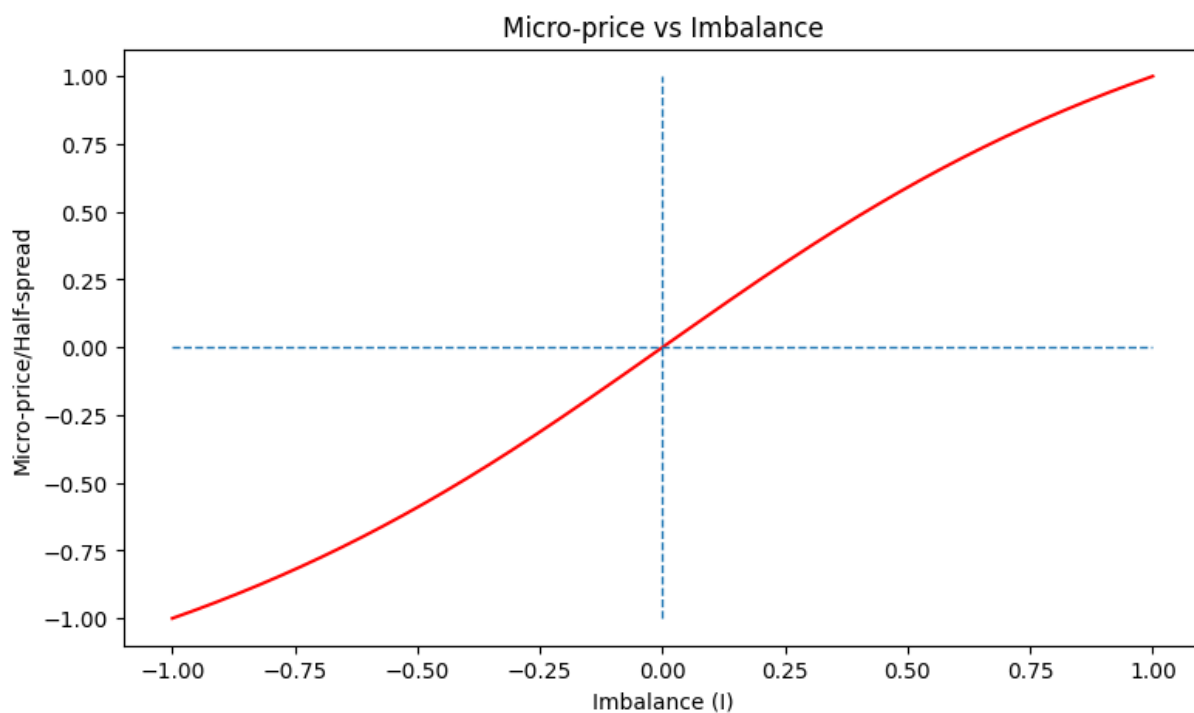

Figure 6: Microprice as a function of the order book imbalance I

- It looks nearly linear!

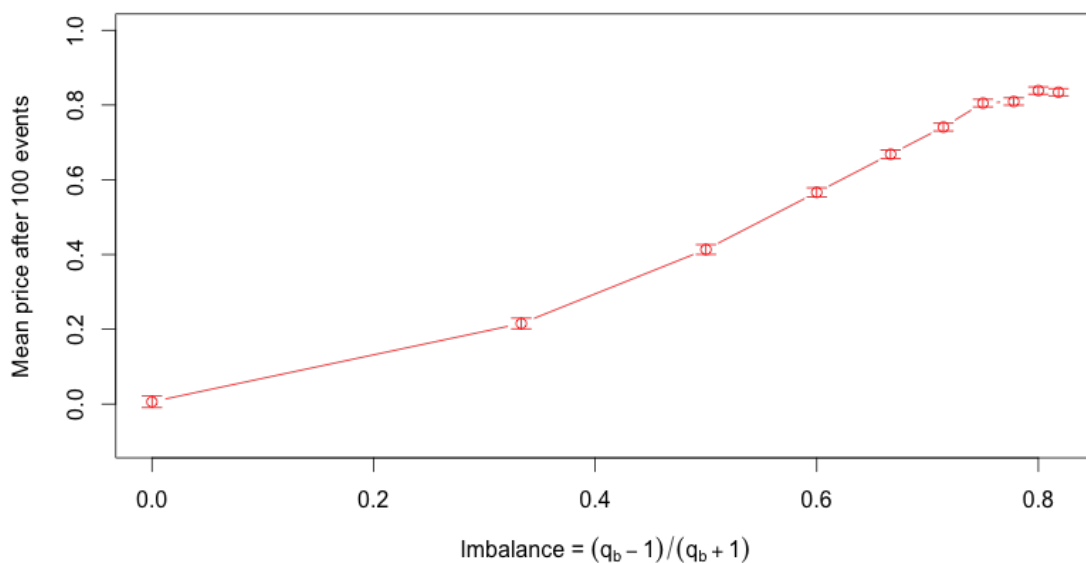## Price signal in the ZI simulation vs book imbalance

Figure 7: With one share at best offer, future price change vs book imbalance
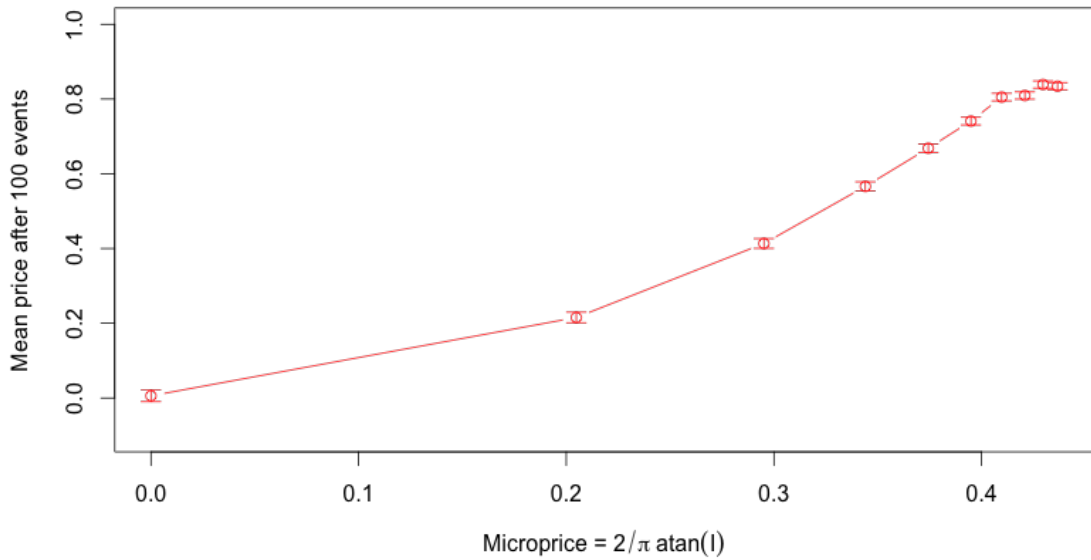
Price signal in the ZI simulation vs microprice

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

Avellaneda, Reed and Stoikov

[Avellaneda, Reed and Stoikov][1] 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)

$$u_{xx} + 2\rho\, u_{xy} + u_{yy} = 0.$$

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.

- $\rho = -1$, 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

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

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][8] 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.

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. ^ Marco Avellaneda, Josh Reed and Sasha Stoikov, Forecasting from level-I quotes in the presence of hidden liquidity, *Algorithmic Finance*, **1**(1) 35-43 (2011).
2. ^ Jean-Philippe Bouchaud, Juius Bonart, Jonathan Donier, and Martin Gould, *Trades Quotes and Prices*, Cambridge University Press (2018).
3. ^ Rama Cont and Adrien de Larrard, Price dynamics in a Markovian limit order market, *SIAM J. Finan. Math.*, **4**(1) 1-25 (2013).
4. ^ Rama Cont and Adrien de Larrard, Order book dynamics in liquid markets: limit theorems and diffusion approximations, *SSRN*, (2011).
5. ^ Rama Cont, Sasha Stoikov and Rishi Talreja, A stochastic model for order book dynamics, *Operations Research*, **58** 549-563 (2010).
6. ^ 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. ^ 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. ^ Ioane Muni Toke and Nakahiro Yoshida, Modelling intensities of order flows in a limit order book, *Quantitative Finance* **17**(5) 683-701 (2017).