PRINCETON UNIVERSITY

ORF 474: High Frequency Trading
Spring 2020
Robert Almgren

Lecture 9a

April 6, 2020

Today

- Begin development of trading signals

  Signals: price dynamics affects trading

  Impact: trading affects price dynamics (after signals)
- Simulation approaches / how to evaluate a signal

  forward correlation

  implementation in trading system

High Frequency Trading

- How to make money from markets:
- Services

  provide technology, data, etc

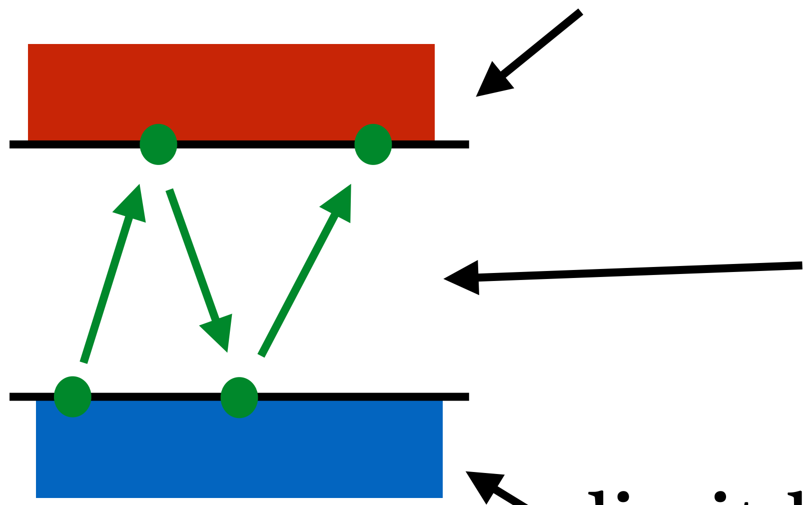
  provide execution services

  provide liquidity (market making, dealing)
- Trading

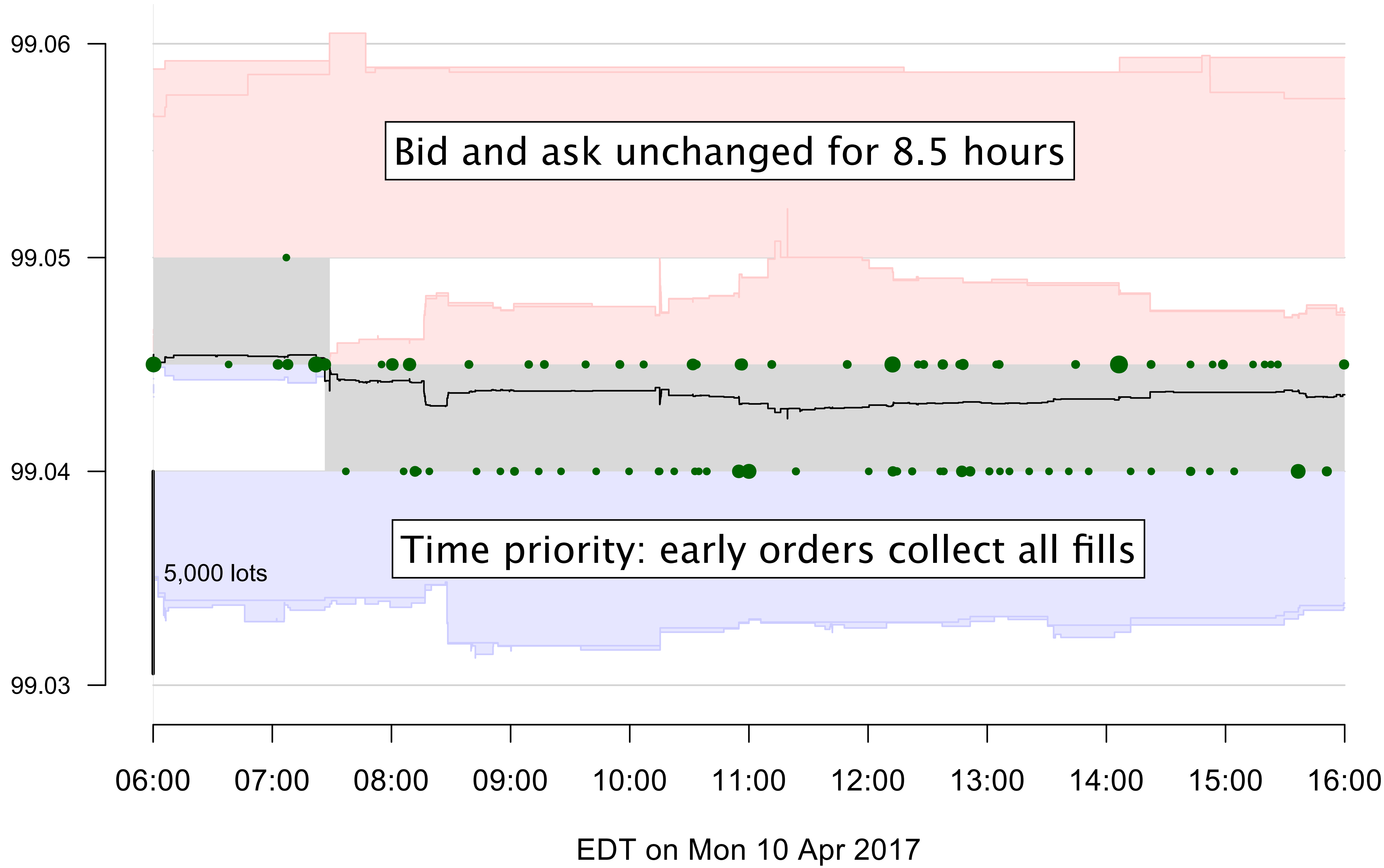
  price forecasting

  taking risk

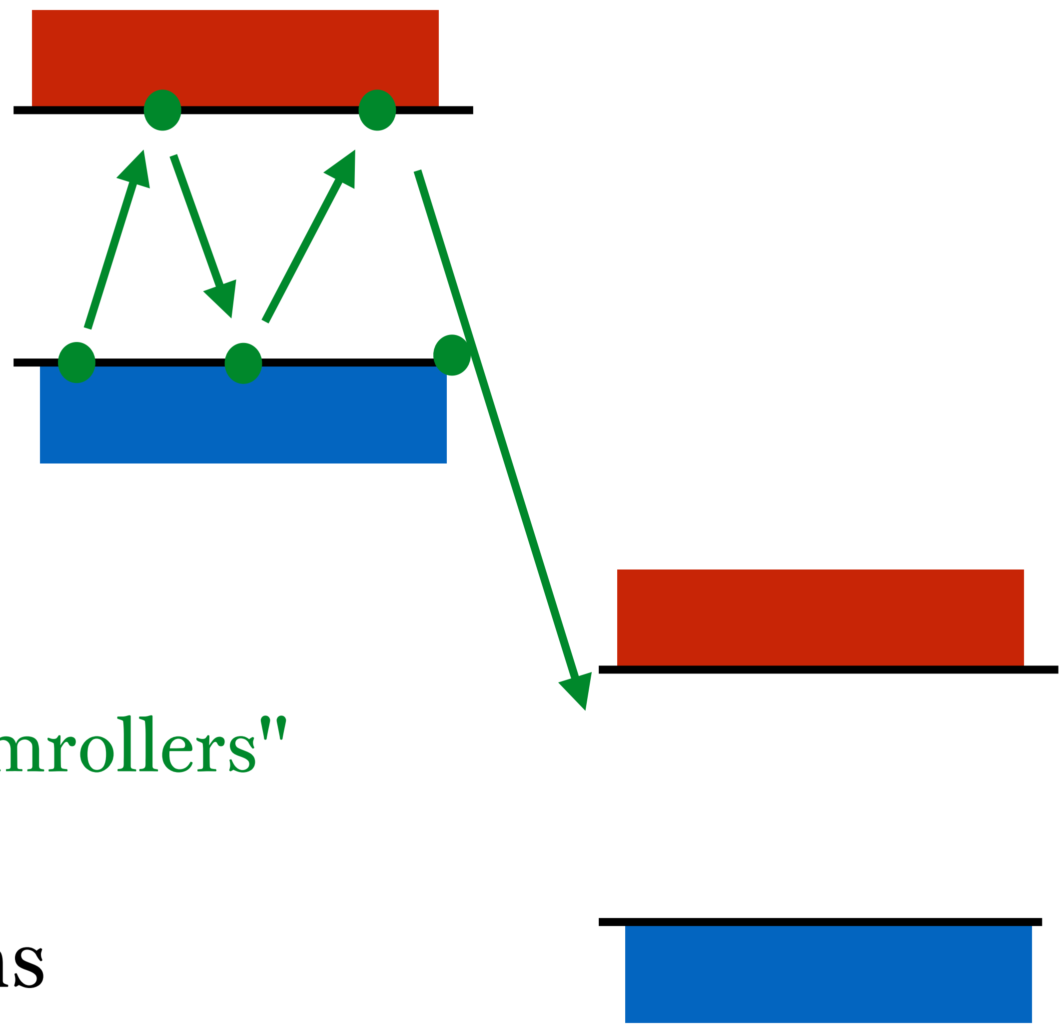
Liquidity provision: market making



BAXM7: June 2017 Canadian Banker's Acceptance (a short term interest rate, like Eurodollar)



Market making



- Problem: trade throughs
  - you are long as price drops
  - "Picking up nickels in front of steamrollers"
  - Competition drives profits down
- Solution: anticipate price motions
  - pull quotes before price moves
  - fast technology is important
- Also: exchanges maintain large ticks
  - market makers are important for liquidity
  - socially useful

Market making profits

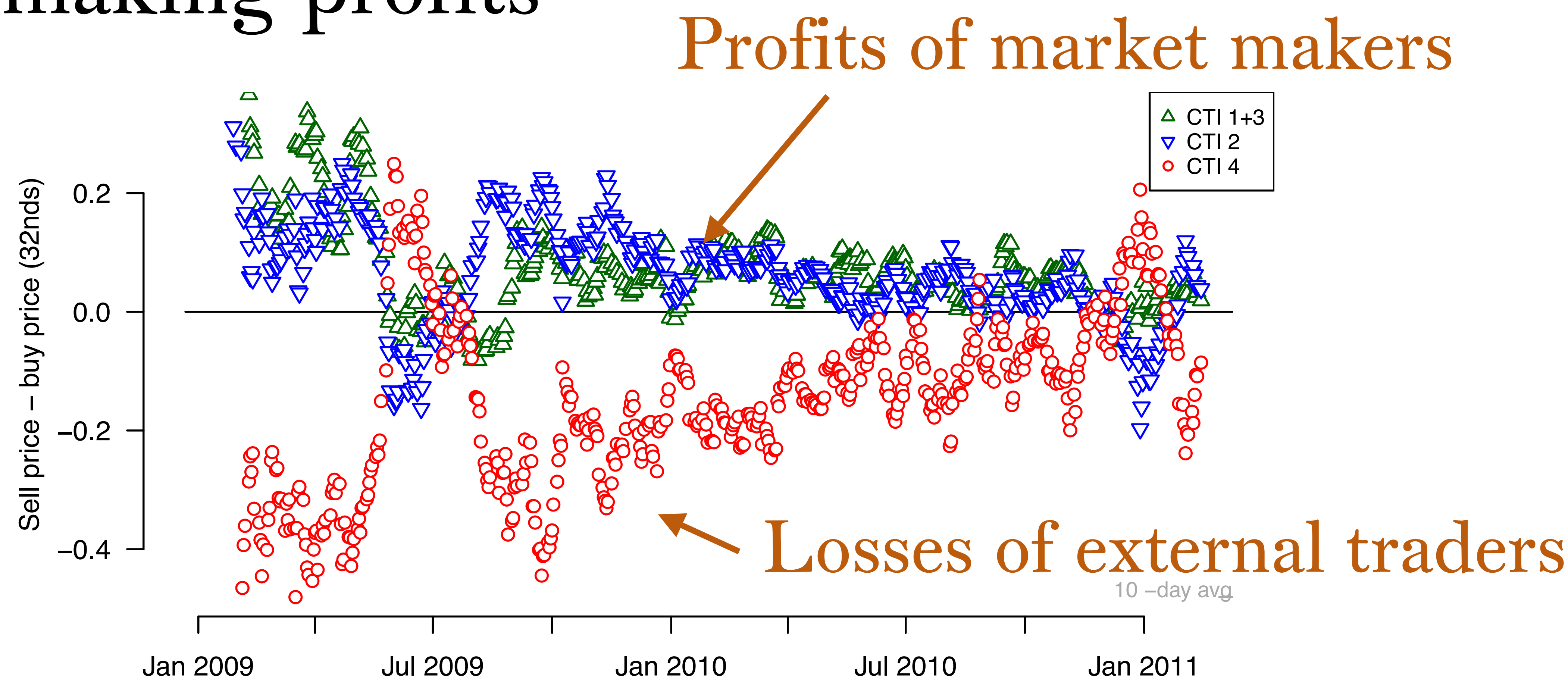
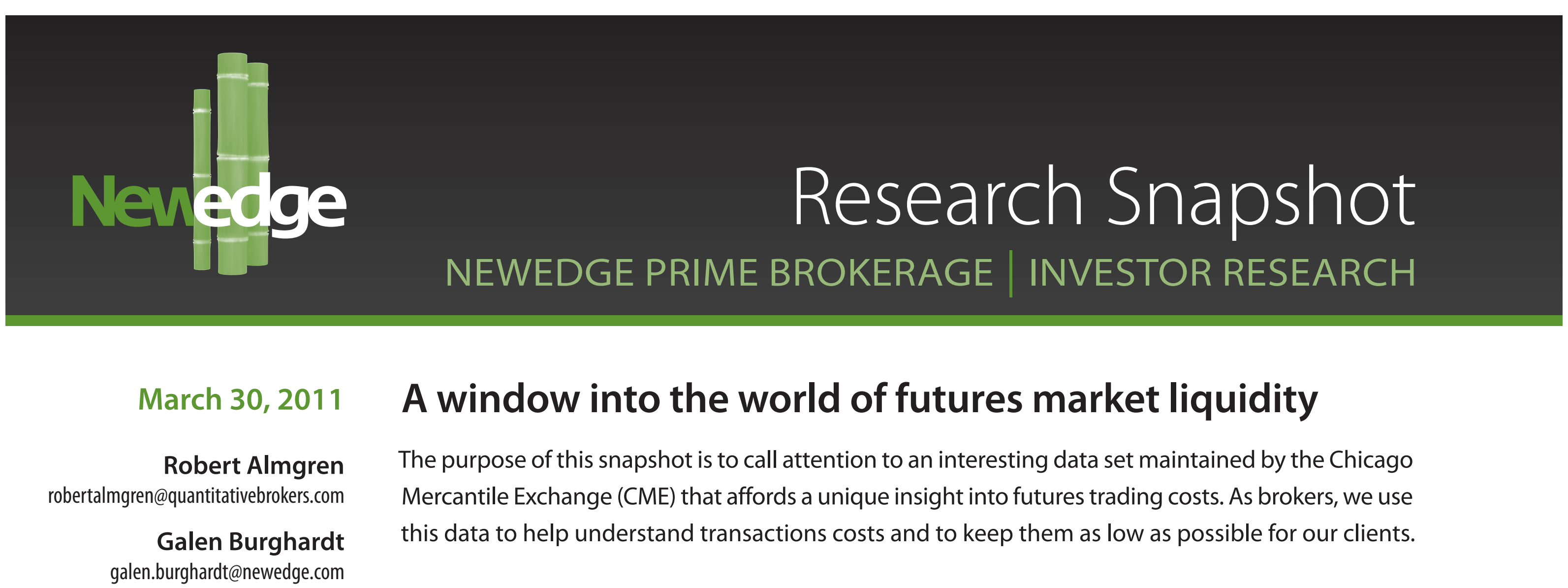


Figure 2: Difference between day's average sell price and day's average buy price, for the 10-year Treasury futures contract. External traders (CTI 4) have been paying an ever-decreasing premium to the "locals" (CTI 1) and the member firms (CTI 2).



Newedge

Research Snapshot

NEWEDGE PRIME BROKERAGE | INVESTOR RESEARCH

March 30, 2011

Robert Almgren
robertalmgren@quantitativebrokers.com

Galen Burghardt
galen.burghardt@newedge.com

A window into the world of futures market liquidity

The purpose of this snapshot is to call attention to an interesting data set maintained by the Chicago Mercantile Exchange (CME) that affords a unique insight into futures trading costs. As brokers, we use this data to help understand transactions costs and to keep them as low as possible for our clients.

CME LDB data set

TABB FORUM
WHERE CAPITAL MARKETS SPEAK

Market Making – Behind the Models

Larry Tabb, TABB Group
30 September 2016

> *Modern market making is misunderstood. While the business model of market makers may have provided easy profits decades ago, today's markets are more complex, efficient, technology-driven, and lightning fast, and market makers no longer have unique access to information. In fact, many liquidity providers prefer not to be market makers, seeking refuge in anonymity, rather than the relatively meager advantages and the somewhat harsh spotlight involved with becoming a transparent and registered market maker. So how do market makers earn a profit?*

Market making is misunderstood. Common wisdom suggests market makers sit on the bid and offer of each product waiting for an investor to take their liquidity. Once acquired, the liquidity is quickly turned over, providing the market maker with a nice and easy profit that very well may be in conflict with original investor. While that may have been the business model of the market makers a decade ago, market making has grown up since a single trader had unique access to proprietary information from exchanges that had garnered a 90% market share.

...

Multiple, competitive markets make it harder and more expensive for market makers to obtain, aggregate, and analyze trading information to provide accurate and competitive prices to investors. In addition, SEC regulations, advances in trading technology, and exchange pricing strategies have enabled non-market makers to trade at comparable prices.

...

While market makers need to invest heavily in connecting to many venues, market making trading strategies also have changed. No longer can market makers just sit on the bid and offer, hoping to profit off the spread. Today, market makers generally have four trading tools/mechanisms: segmentation, fair value analytics, single product supply and demand, and multi-product hedging and correlation strategies.

https://tabbforum.com/opinions/market-making-behind-the-models

Liquidity provision: dealing

- Example:
  Asset manager calls bank desk to execute large trade
  hundreds of names, 10s of $MM value
- Value add: ability to commit bank's capital
- Risks:
  - adverse information
  - liquidity costs to unwind position
  - volatility risk while holding position
- Manage risks by knowing counterparty
  - little use of price forecast

Execution services

- Agency execution: no capital at risk
- Value add
  - expertise
  - technology
  - customer service
- Improve by giving better execution
- Most important: price forecasts

Trading

- Take positions in market to make profits with your own money or others' money
- Other people's money
  - providing a service
    - decent performance
    - uncorrelated
  - still need return
  - you are probably compensated on P&L
- Your money
  - pure alpha (return) or whatever you prefer
  - most important is price prediction

Investing is providing service and products

Our Company and Sites

BLACKROCK

Products | Insights | Retirement | Resources

BY TYPE

Mutual Funds

Fixed Income Funds

Equity Funds

Multi-Asset Funds

iShares ETFs

Closed End Funds

MORE PRODUCTS & RESOURCES

Managed Accounts

Target Date Funds

CoRI Funds

Variable Insurance Funds

529 College Saving Plans

Alternative Investments

Cash and Liquidity Products

Closed End Funds Resources

Model Portfolios

Conclusion

- There are many ways to make money in markets

  liquidity provision and capital commitment

  services

  trading
- They require many skills and assets

  market understanding

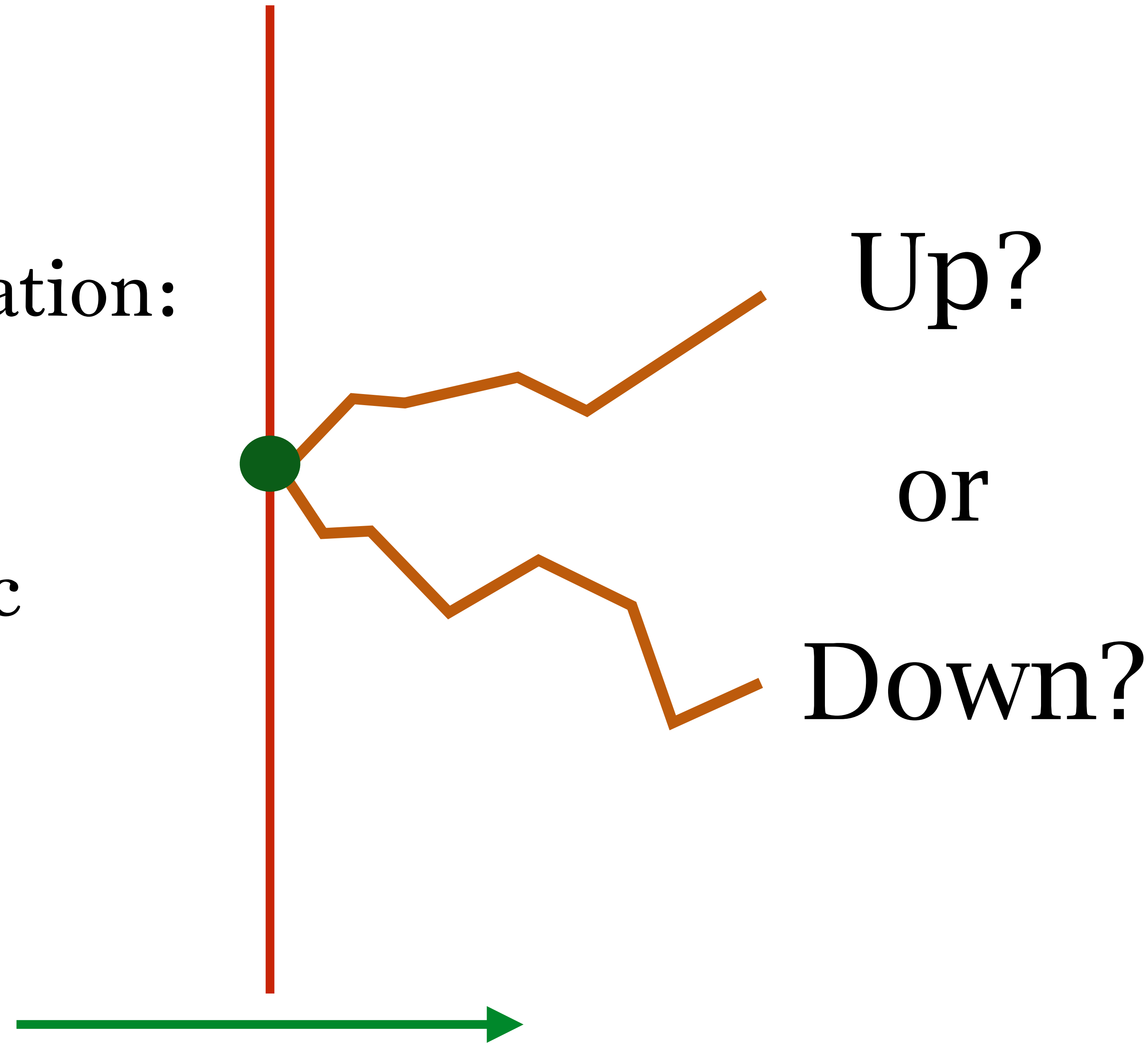
  customer relations

  reliable performance

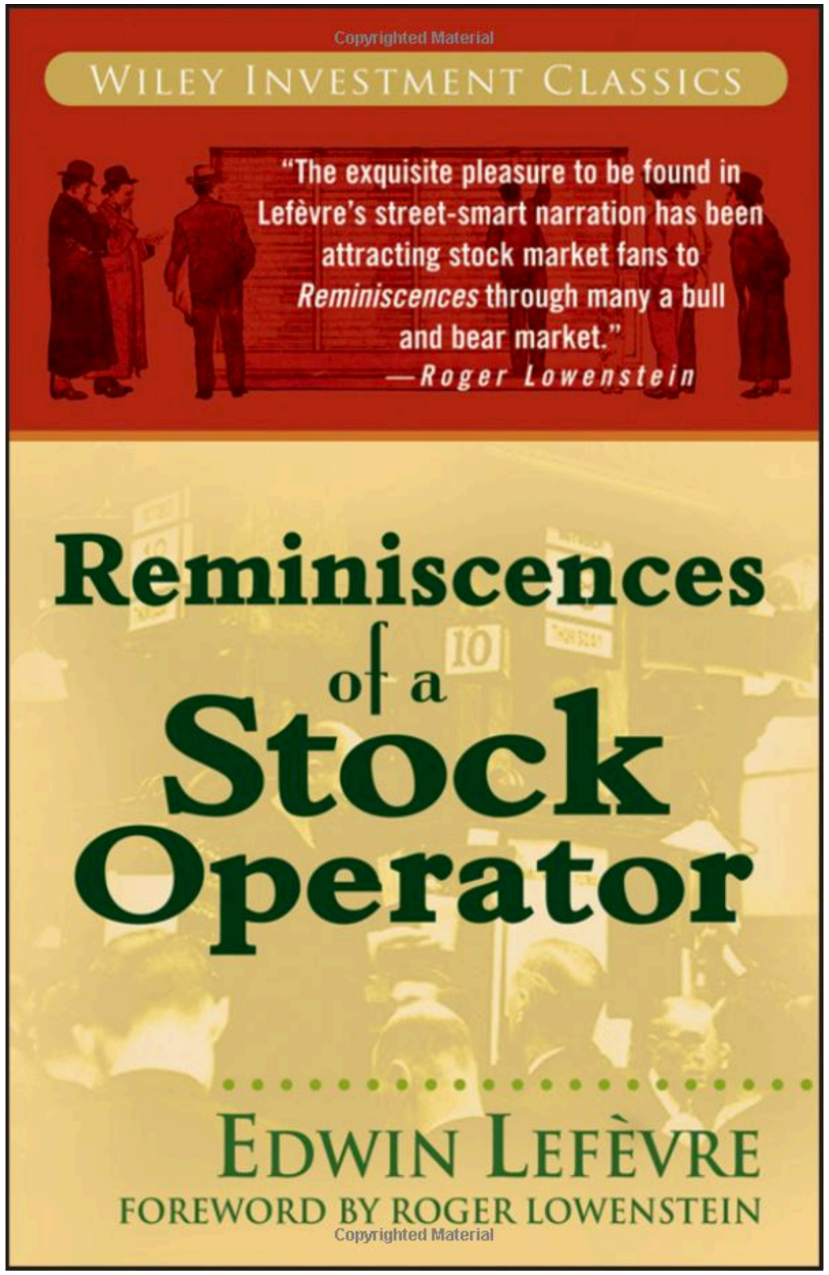
  high-performance technology
- All require some form of price prediction

Price prediction

All possible information:
- price series of this and other assets
- trades, volume, etc
- external info



Past Now Future



I have heard of people who amuse themselves conducting imaginary operations in the stock market to prove with imaginary dollars how right they are. Sometimes these ghost gamblers make millions. It is very easy to be a plunger that way. It is like the old story of the man who was going to fight a duel the next day.

His second asked him, "Are you a good shot?"

"Well," said the duelist, "I can snap the stem of a wine-glass at twenty paces," and he looked modest.

"That's all very well," said the unimpressed second. "But can you snap the stem of the wineglass while the wineglass is pointing a loaded pistol straight at your heart?"

External information

BRIDGEWATER

Bridgewater Associates is focused on understanding how the world works. By having the deepest possible understanding of the global economy and financial markets, and translating that understanding into great portfolios and strategic partnerships with institutional clients, we've built a distinct track record of success. We've done this for more than 40 years by having great people operate in a culture of radical truth and radical transparency. Today, we manage about $160 billion for approximately 350 of the largest and most sophisticated global institutional clients including public and corporate pension funds, university endowments, charitable foundations, supranational agencies, sovereign wealth funds, and central banks.

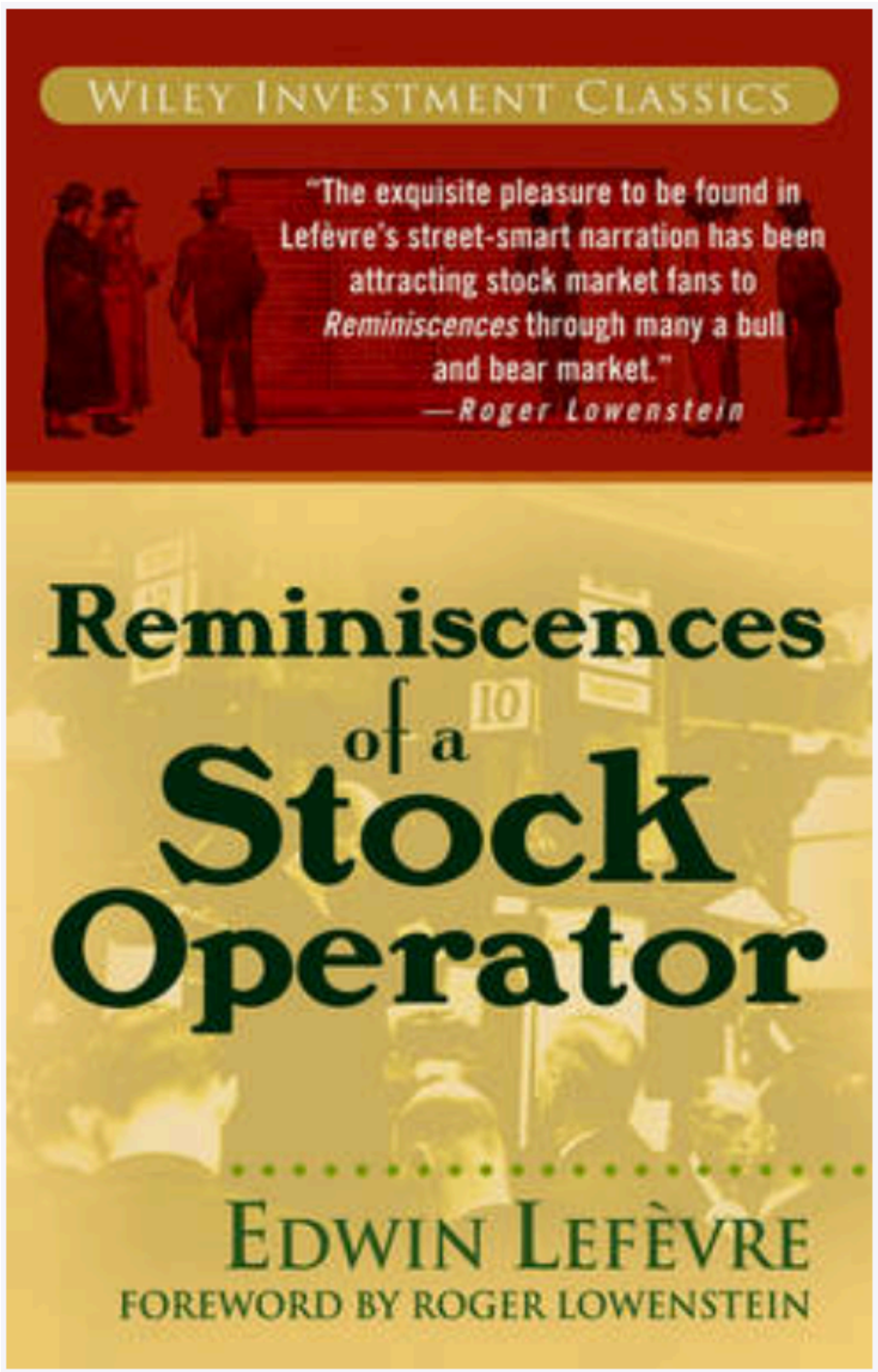
External information sources

- Company and industry information (stocks)
- Macroeconomic views (interest rates, FX)
- Weather forecasts (agricultural, energy)
- Inventory state (oil, etc)
- Parking lot density (retail stocks)
- Text processing
  real-time news feeds
  SEC filings

Internal information sources: market data

- Price histories
  - this asset
  - other futures or other stocks
  - futures → underlying, or vice versa
  - options
- Trade volume dynamics
- Size of individual trades
- Quote dynamics: cancellations, etc
- Intuition: all relevant information is already reflected in market data

Market information



You watch the market -- that is, the course of prices as recorded by the tape with one object: to determine the direction -- that is, the price tendency. Prices, we know, will move either up or down according to the resistance they encounter. For purposes of easy explanation we will say that prices, like everything else, move along the line of least resistance. They will do whatever comes easiest, therefore they will go up if there is less resistance to an advance than to a decline; and vice versa.

Price prediction is hard

- Easy to predict nontradeable numbers
  - volume, volatility, etc
  - physical variables: weather, etc
  - prices that are not tradeable
- Efficient market theory says that price prediction is extremely hard
  - prediction changes the price itself
- Need to have some idea what your signal is

Price prediction is separate from execution

- Once you have a price forecast, how do you execute?
- Separate problem from signal identification
- Different for prop trading vs agency execution
  - Prop trading
    - Must execute round-trip trades
    - But can choose time
  - Agency execution
    - only shifting time of execution
    - but cannot choose when or what to execute
- Execution can be important part of P&L
- Ultimate test is overall performance of strategy
  - Need framework for backtesting

Sources of price prediction from market data

- Reversion from sudden moves
- Less liquid asset follows more liquid asset
- Identifying large traders with directional goals

Almost every signal is one of these types

Time scales

- Short-term

  microseconds to milliseconds to seconds

  need good tick data and good technology
- Medium-term

  minutes to hours

  can use coarser data (even minute bars)

  relationships between different assets, including options
- Long-term

  daily and longer

  fundamental information (stocks, etc)

  macro-economic views

Examples

- Bid-ask quote imbalance

  imbalance indicates finer view of "true price"
- Trade direction imbalance

  indicates preferential trade directions
- "Sweep" signal for rapid reversion

  sharp motions are often overreach and come back
- Pairs trading between two or assets

  maintain equilibrium relationship
- Trade size

  specific sizes are used by uninformed traders

Signal development and evaluation

1. Propose idea
   plausibility tests
2. Statistical tests on historical market data
   nonzero correlation with future price movement
3. Rerun executed orders, or trading strategy
   show improvement in slippage or in P&L

Market simulator

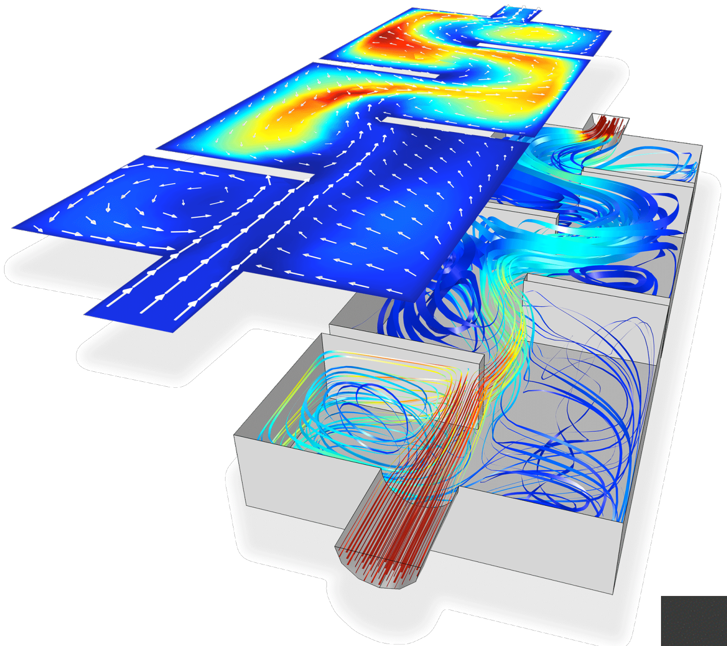
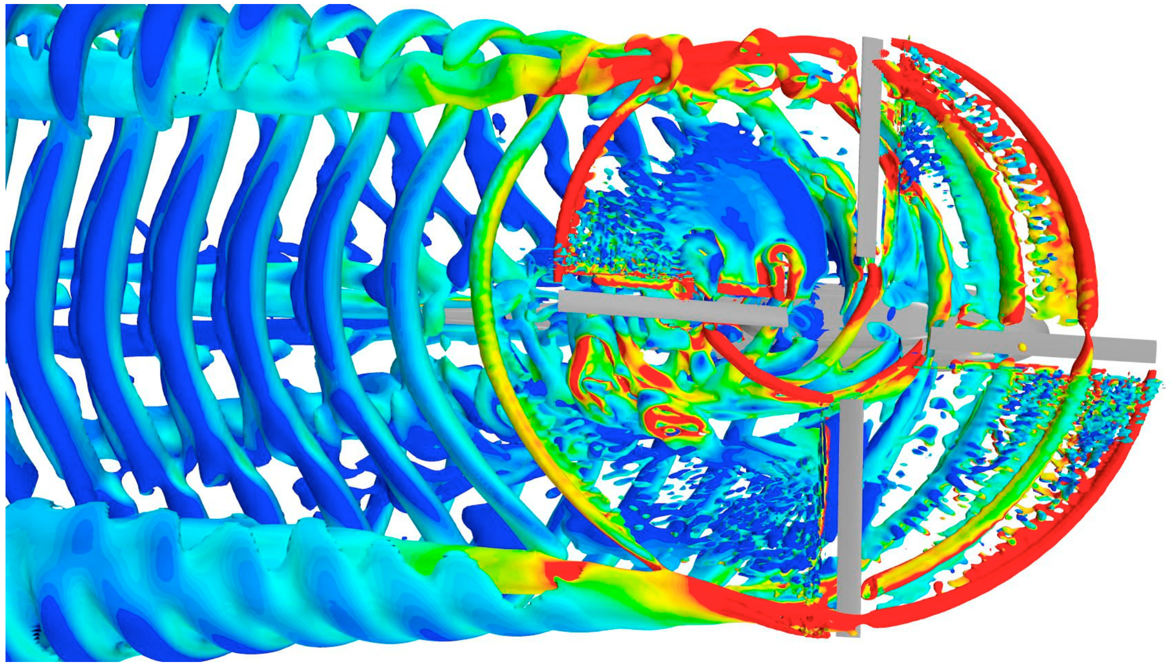
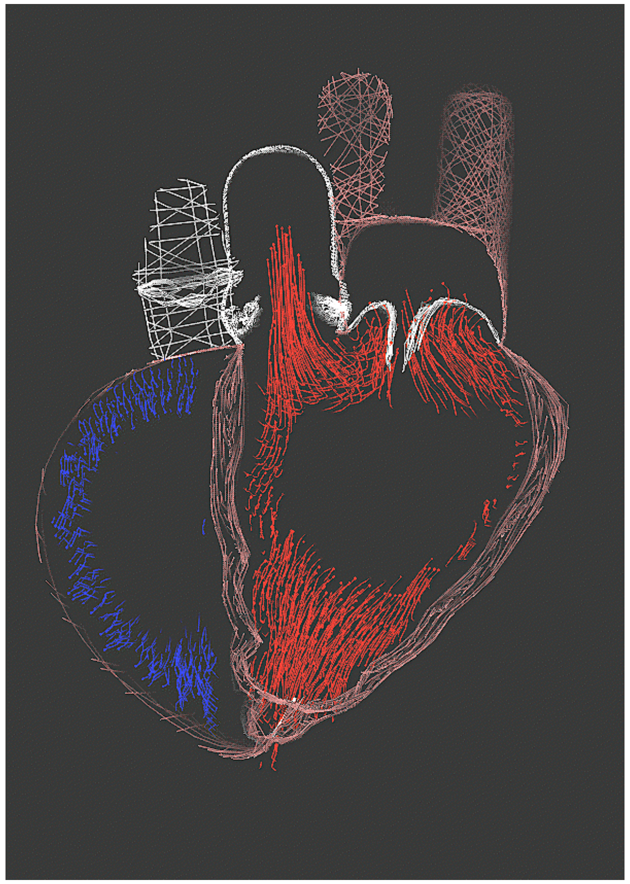
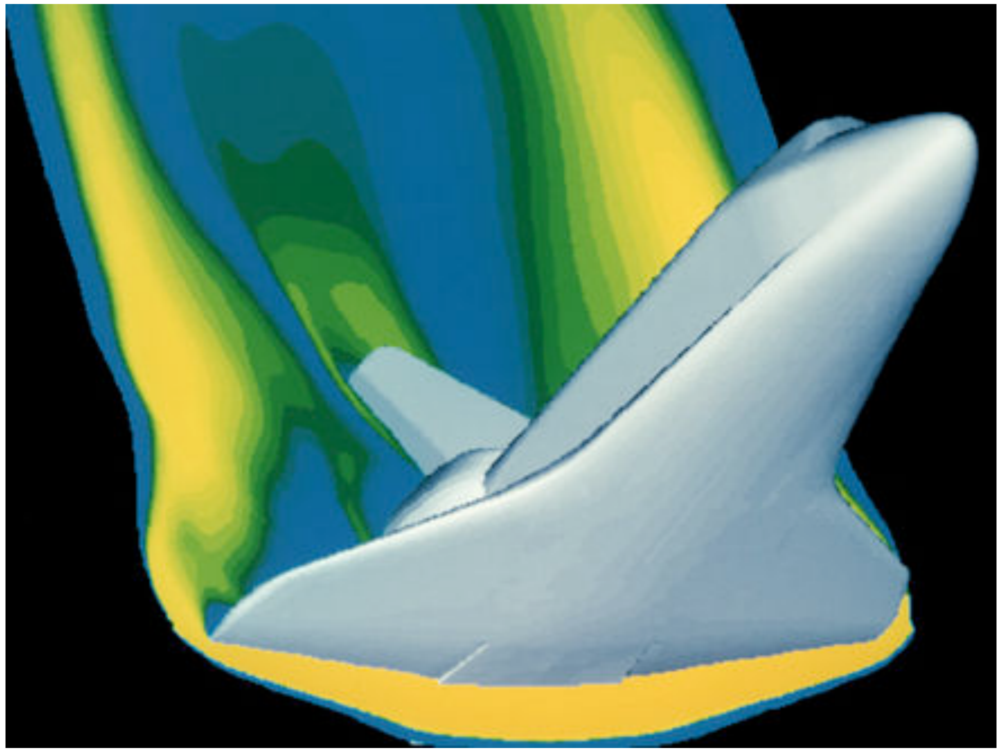
- Realistic reproduction of market (in some sense)
- Use to test various aspects of trading systems:

  Basic functionality

  Statistical techniques (volatility)

  Response to signals

Computational fluid dynamics

Computational fluid dynamics

- Complete simulation is impossible
- Discretize to capture key features:
- Conservation of mass, momentum, etc
- Positivity of density, etc
- Vortex dynamics
- Chemical reactions
- 2-D, 3-D, axisymmetric, etc
- Nonlocal effects (incompressible flow)

Simulation/backtesting techniques

- Stochastic simulation
  - Calibrate stochastic model to market data
  - Zero-intelligence order book dynamics
  - Calibrated Markov model to order book dynamics
- Replay actual market data

- Stochastic simulation
  - Pro: can do arbitrarily many sample paths
  - Con: only partially realistic
- Replay actual market data
  - Pro: finite amount of actual history
  - Con: accurate market dynamics, no market impact

Pure stochastic simulation

Machine learning for trading

risk.net October 2017

In multiperiod trading with realistic market impact, determining the dynamic trading strategy that optimises the expected utility of final wealth can be difficult. Gordon Ritter shows that, with an appropriate choice of reward function, reinforcement learning techniques (specifically Q-learning) can successfully handle the risk-averse case

For this example, assume there exists a tradable security with a strictly positive price process $p_t > 0$. (This 'security' could itself be a portfolio of other securities, such as an exchange-traded fund or a hedged relative-value trade.) Further, suppose there is some 'equilibrium price' p_{e} such that $x_t = \log(p_t / p_{\mathrm{e}})$ has dynamics:

$$\mathrm{d}x_t = -\lambda x_t + \sigma \xi_t \tag{16}$$

where $\xi_t \sim N(0, 1)$, and ξ_t and ξ_s are independent when $t \neq s$. This means that p_t tends to revert to its long-run equilibrium level p_{e}, with mean-reversion rate λ, and is a standard discretisation of the Ornstein-Uhlenbeck process. For this exercise, the parameters of the dynamics (16) were taken to be $\lambda = \log(2)/H$, where $H = 5$ is the half-life, $\sigma = 0.1$, and the equilibrium price is $p_{\mathrm{e}} = 50$.

Gordon Ritter is a senior portfolio manager at GSA Capital Partners, and adjunct professor at Courant Institute at NYU, Baruch College (CUNY) and Department of Statistics Rutgers University. Email: ritter@post.harvard.edu.

A major drawback of the procedure we have presented here is it requires a large number of training steps (a few million for the problem we presented). There are, of course, financial data sets with millions of time steps (eg, high-frequency data sampled once per second for several years), but a different approach is needed in other cases. Even in high-frequency examples, one may not wish to use several years' worth of data to train the model.

Fortunately, a simulation-based approach presents an attractive resolution to these issues. In other words, we propose a multistep training procedure: (1) posit a reasonably parsimonious stochastic process model for asset returns with relatively few parameters; (2) estimate the parameters of the model from market data, ensuring reasonably small confidence intervals for the parameter estimates; (3) use the model to simulate a much larger data set than the real world presents; and (4) train the reinforcement learning system on the simulated data.

Zero-intelligence model

QUANTITATIVE FINANCE VOLUME 3 (2003) 481–514 RESEARCH PAPER
INSTITUTE OF PHYSICS PUBLISHING quant.iop.org

Statistical theory of the continuous double auction

Eric Smith[1], J Doyne Farmer, László Gillemot and Supriya Krishnamurthy

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

E-mail: desmith@santafe.edu

Received 30 October 2002, in final form 16 July 2003
Published 18 September 2003
Online at stacks.iop.org/Quant/3/481

Abstract
Most modern financial markets use a continuous double auction mechanism to store and match orders and facilitate trading. In this paper we develop a microscopic dynamical statistical model for the continuous double auction under the assumption of IID random order flow, and analyse it using simulation, dimensional analysis, and theoretical tools based on mean field approximations. The model makes testable predictions for basic properties of markets, such as price volatility, the depth of stored supply and demand versus price, the bid–ask spread, the price impact function, and the time and probability of filling orders. These predictions are based on properties of order flow and the limit order book, such as share volume of market and limit orders, cancellations, typical order size, and tick size. Because these quantities can all be measured directly there are no free parameters. We show that the order size, which can be cast as a non-dimensional granularity parameter, is in most cases a more significant determinant of market behaviour than tick size. We also provide an explanation for the observed highly concave nature of the price impact function. On a broader level, this work suggests how stochastic models based on zero intelligence agents may be useful to probe the structure of market institutions. Like the model of perfect rationality, a stochastic zero intelligence model can be used to make strong predictions based on a compact set of assumptions, even if these assumptions are not fully believable.

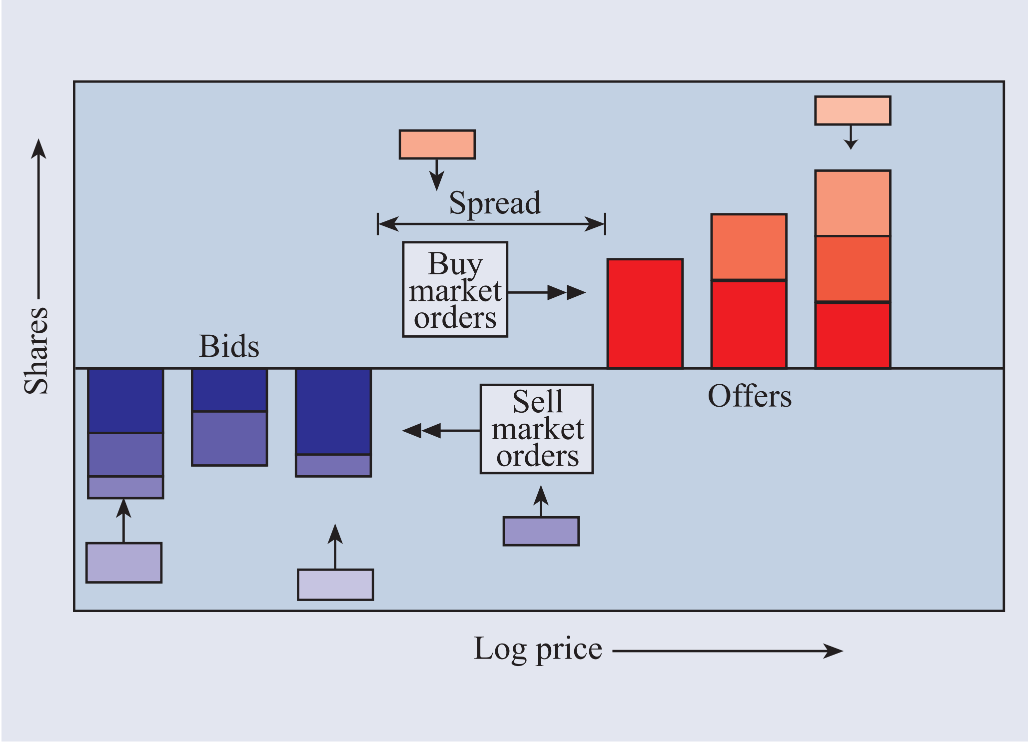


Figure 1. A schematic illustration of the continuous double auction mechanism and our model of it. Limit orders are stored in the limit order book. We adopt the arbitrary convention that buy orders are negative and sell orders are positive. As a market order arrives, it has transactions with limit orders of the opposite sign, in order of price (first) and time of arrival (second). The best quotes at prices $a(t)$ or $b(t)$ move whenever an incoming market order has sufficient size to fully deplete the stored volume at $a(t)$ or $b(t)$. Our model assumes that market order arrival, limit order arrival, and limit order cancellation follow a Poisson process. New offers (sell limit orders) can be placed at any price greater than the best bid, and are shown here as 'raining down' on the price axis. Similarly, new bids (buy limit orders) can be placed at any price less than the best offer. Bids and offers that fall inside the spread become the new best bids and offers. All prices in this model are logarithmic.

Poisson process for orders & cancellations

- New limit orders: α unit time per unit price

  Buy orders: from a(t) to -Inf

  Sell orders: from b(t) to +Inf

- Limit order cancellation: δ per unit time per order
- Market orders: μ per unit time

Price can be discrete or continuous

- Away from bid-ask, get steady state order distribution
- Limit sells at p<a(t) or limit buys at p>b(t) decrease spread
- Market orders increase spread

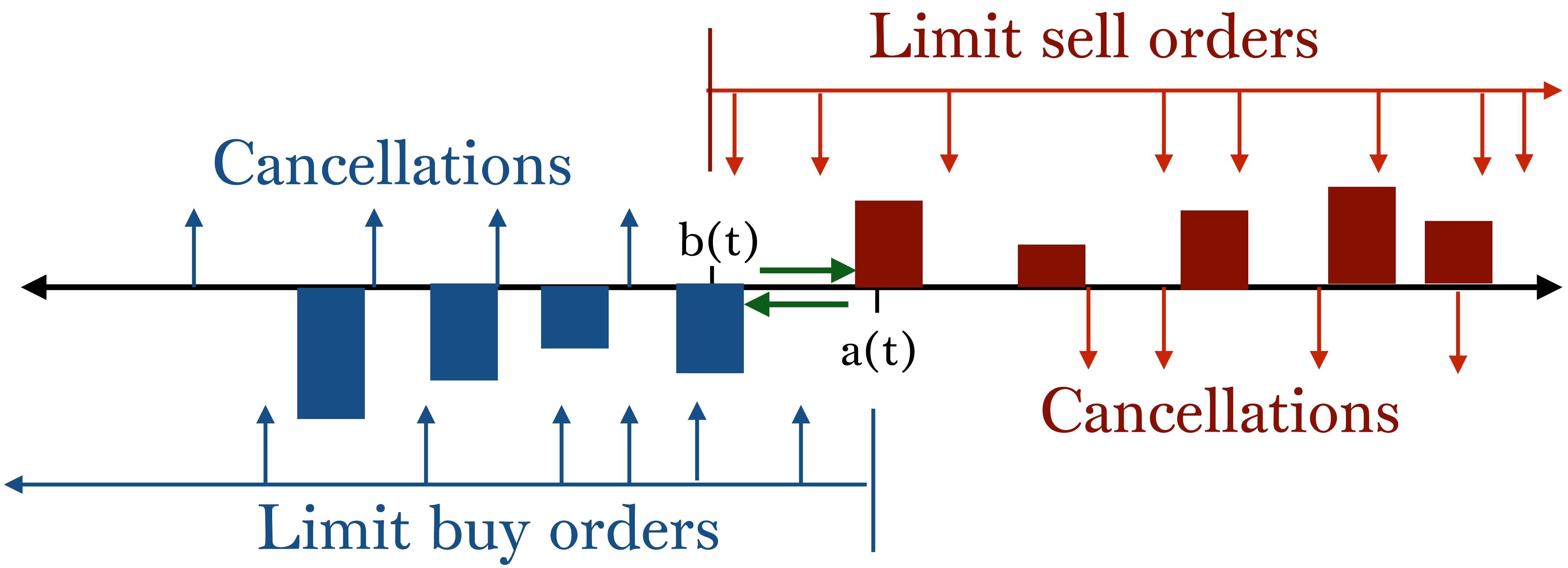


Table 1. The five parameters that characterize this model. α, μ and δ are order flow rates, and $\mathrm{d}p$ and σ are discreteness parameters.

Parameter	Description	Dimensions
α	Limit order rate	Shares/(price time)
μ	Market order rate	Shares/time
δ	Order cancellation rate	1/time
$\mathrm{d}p$	Tick size	Price
σ	Characteristic order size	Shares

Table 2. Important characteristic scales and non-dimensional quantities. We summarize the characteristic share size, price and times defined by the order flow rates, as well as the two non-dimensional scale parameters $\mathrm{d}p/p_c$ and ϵ that characterize the effect of finite tick size and order size. Dimensional analysis makes it clear that all the properties of the limit order book can be characterized in terms of functions of these two parameters.

Parameter	Description	Expression
N_c	Characteristic number of shares	$\mu/2\delta$
p_c	Characteristic price interval	$\mu/2\alpha$
t_c	Characteristic time	$1/\delta$
$\mathrm{d}p/p_c$	Non-dimensional tick size	$2\alpha\,\mathrm{d}p/\mu$
ϵ	Non-dimensional order size	$2\delta\sigma/\mu$

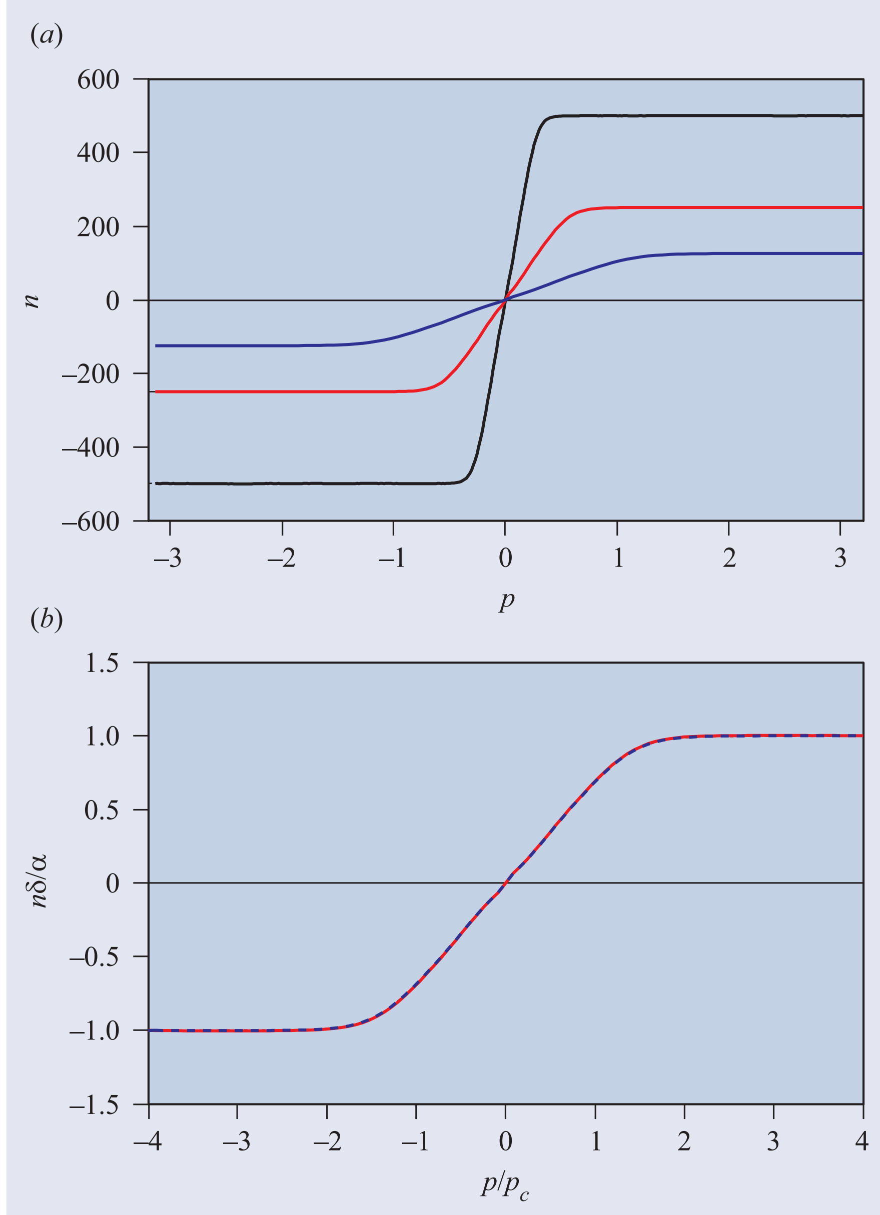
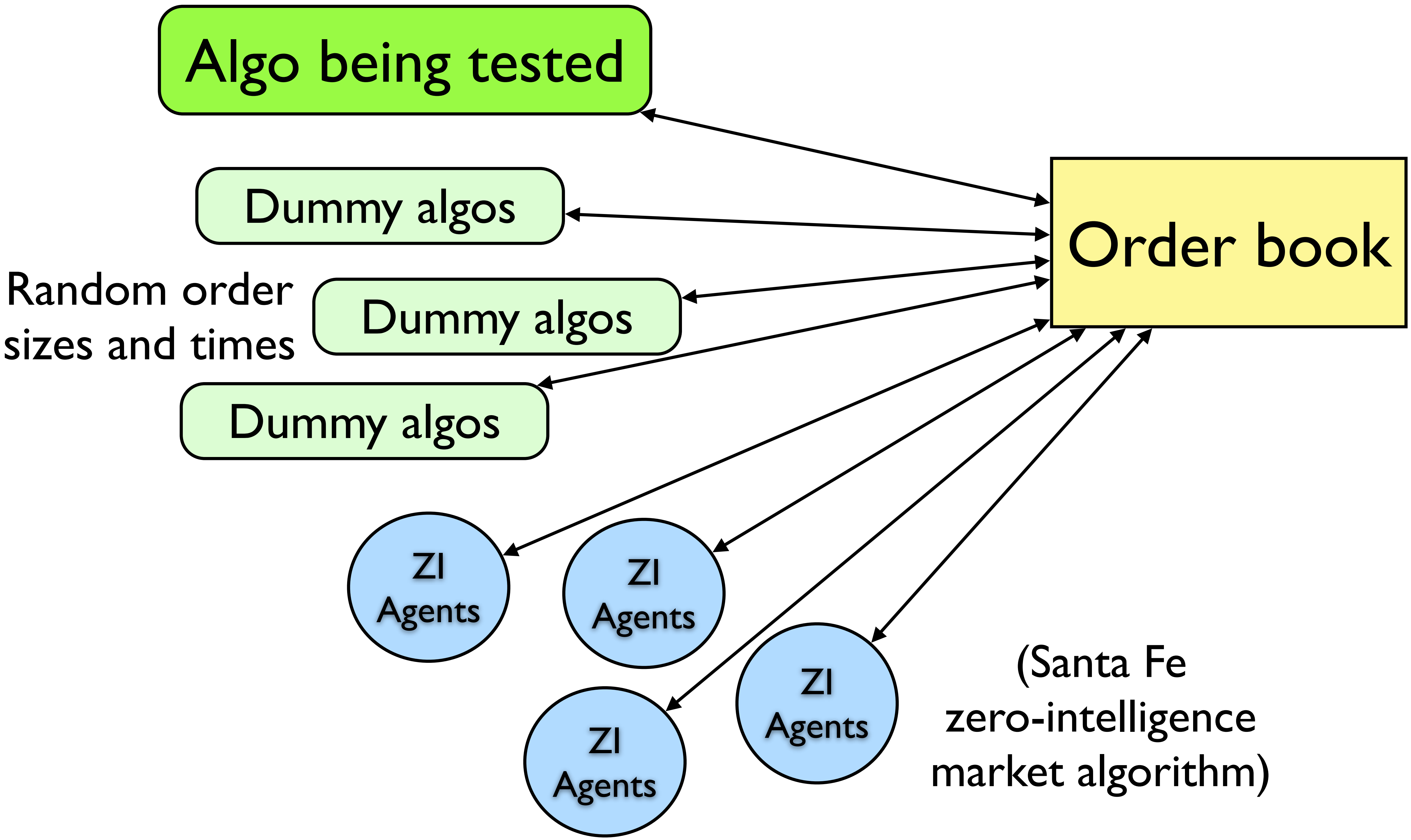


Figure 2. The usefulness of non-dimensional units. (*a*) We show the average depth profile for three different parameter sets. The parameters $\alpha = 0.5$, $\sigma = 1$, and $\mathrm{d}p = 0$ are held constant, while δ and μ are varied. The curve types are: black: $\delta = 0.001$, $\mu = 0.2$; red: $\delta = 0.002$, $\mu = 0.4$; and blue: $\delta = 0.004$, $\mu = 0.8$. (*b*) The same, but plotted in non-dimensional units. All three curves collapse on top of each other. The horizontal axis has units of *price*, and so has non-dimensional units $\hat{p} = p/p_c = 2\alpha p/\mu$. The vertical axis has units of n shares/price, and so has non-dimensional units $\hat{n} = np_c/N_c = n\delta/\alpha$. Because we have chosen the parameters to keep the non-dimensional order size ϵ constant, the collapse is perfect. Varying the tick size has little effect on the results other than making them discrete.

Use of zero-intelligence model



Pros and cons of Santa Fe model

- Pros:
  - Beautifully simple
  - Useful ingredient in more complex simulation
  - Highlights simplicity of order book dynamics
  - Includes simple version of market impact
- Cons:
  - No realistic price dynamics
  - No actual traders

Simulating and Analyzing Order Book Data: The Queue-Reactive Model

Weibing HUANG, Charles-Albert LEHALLE, and Mathieu ROSENBAUM

Journal of the American Statistical Association
March 2015, Vol. 110, No. 509, Applications and Case Studies

Through the analysis of a dataset of ultra high frequency order book updates, we introduce a model which accommodates the empirical properties of the full order book together with the stylized facts of lower frequency financial data. To do so, we split the time interval of interest into periods in which a well chosen reference price, typically the midprice, remains constant. Within these periods, we view the limit order book as a Markov queuing system. Indeed, we assume that the intensities of the order flows only depend on the current state of the order book. We establish the limiting behavior of this model and estimate its parameters from market data. Then, to design a relevant model for the whole period of interest, we use a stochastic mechanism that allows to switch from one period of constant reference price to another. Beyond enabling to reproduce accurately the behavior of market data, we show that our framework can be very useful for practitioners, notably as a market simulator or as a tool for the transaction cost analysis of complex trading algorithms.

KEY WORDS: Ergodic properties; Execution probability; High frequency data; Jump Markov process; Limit order book; Mechanical volatility; Market impact; Market microstructure; Market simulator; Queuing model; Transaction costs analysis; Volatility.

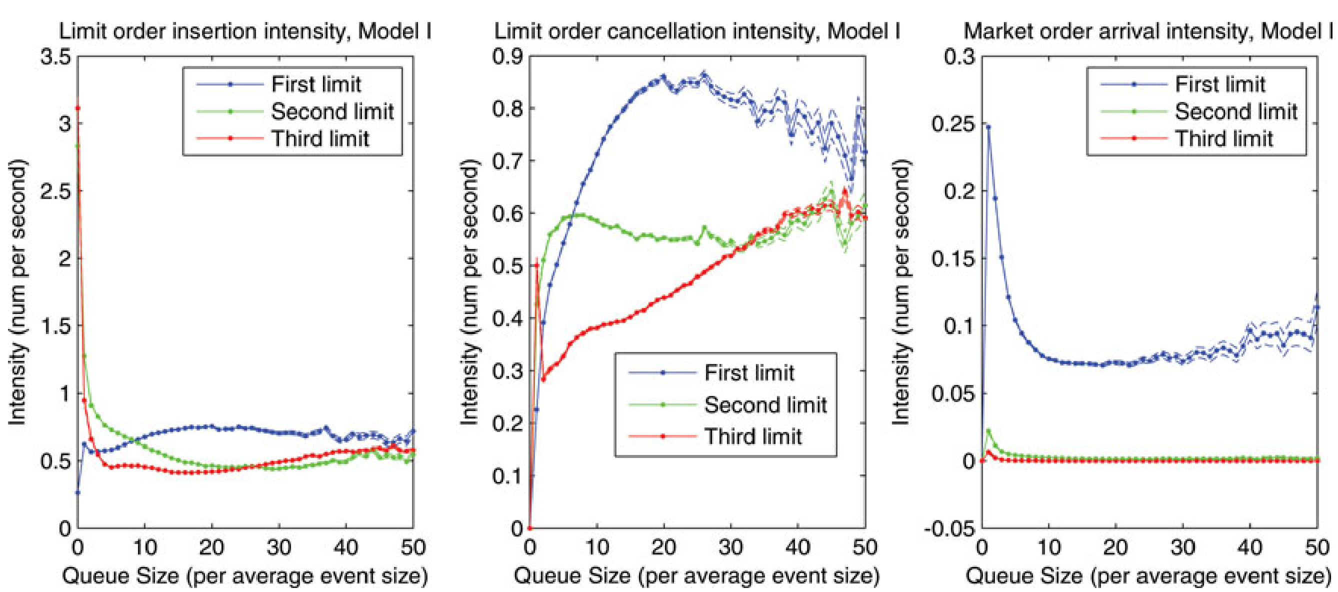


Figure 2. Intensities at $Q_{\pm i}$, $i = 1, 2, 3$, France Telecom.

4. CONCLUSION AND PERSPECTIVES

In this work, we have modeled market participants intelligence through their average behaviors toward various states of the LOB. This enabled us to analyze the different order flows and to design a suitable market simulator for practitioners, allowing notably to investigate the transaction costs of complex trading strategies. To our knowledge, our model is the first one where such pre-trade cost analysis is possible in a simple and efficient way.

Another important public information, the historical order flow, is not considered in this approach. Market order flows have been shown to be autocorrelated in several empirical studies (see, e.g., Toth et al. 2011b). Thus, adding such feature in our framework would probably be relevant. Another possible direction for future research would be to explain the shape of the estimated intensity functions in a more sophisticated way. For example, it would be interesting to design some agent based model where these repetitive patterns of the LOB dynamics would be reproduced, providing an even better understanding of the nature of these intensity curves.

Computational market simulation

- Complete simulation is impossible
- (Human reaction is very complicated)

Key features to include:

queue position and match algorithms

price movement

Features to neglect for simplicity:

market impact

(Literature on agent-based markets)

Market simulator based on real data

- Tool for developing and testing execution algorithms for interest rate products.
- Capture essential features of main markets:
    - matching algorithms and passive fill probabilities
    - short term pricing signals
- Will have limitations -- useful anyway
- Does not embody model of market impact
- The one most natural way to build a simulator

Original idea

The PLAT Project has developed a trading simulation that merges automated clients with real-time, real-world stock market data. This simulation has been used for three competitions.

The Penn-Lehman Automated Trading Project

Michael Kearns and Luis Ortiz, *University of Pennsylvania*

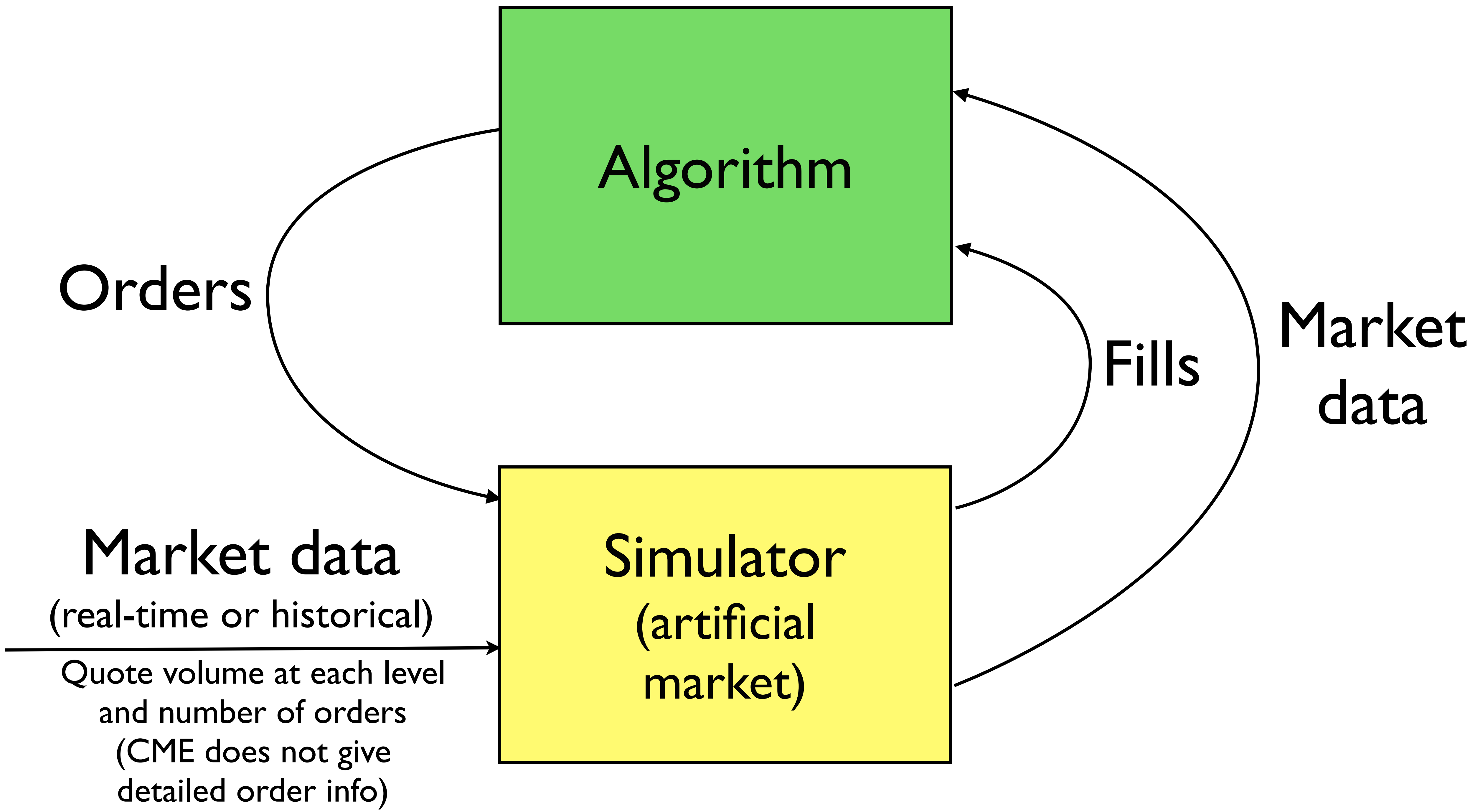
The Penn-Lehman Automated Trading Project is a broad investigation of algorithms and strategies for automated trading in financial markets. The PLAT Project's centerpiece is the Penn Exchange Simulator (PXS), a software simulator for automated stock trading that merges automated client orders for shares with real-world, real-time order data. PXS automatically computes client profits and losses, volumes traded, simulator and external prices, and other quantities of interest. To test the effectiveness of PXS and of various trading strategies, we've held three formal competitions between automated clients.

We also actively use PXS as a platform for developing novel, principled automated trading strategies (clients). The real-data, real-time nature of PXS lets us examine computationally intensive, high-frequency, high-volume trading strategies (although this last property always presents the challenges of estimating the *market impact*—the effect on prices). We're particularly interested in developing clients that make predictive use of limit order book data, including those using statistical modeling and machine learning. We hope that, over time, the project will generate a library of clients with varying features (trading strategy, volume, frequency, and so on) that can serve to create realistic simulations with known properties.

1094-7167/03/$17.00 © 2003 IEEE
Published by the IEEE Computer Society
IEEE INTELLIGENT SYSTEMS

2003

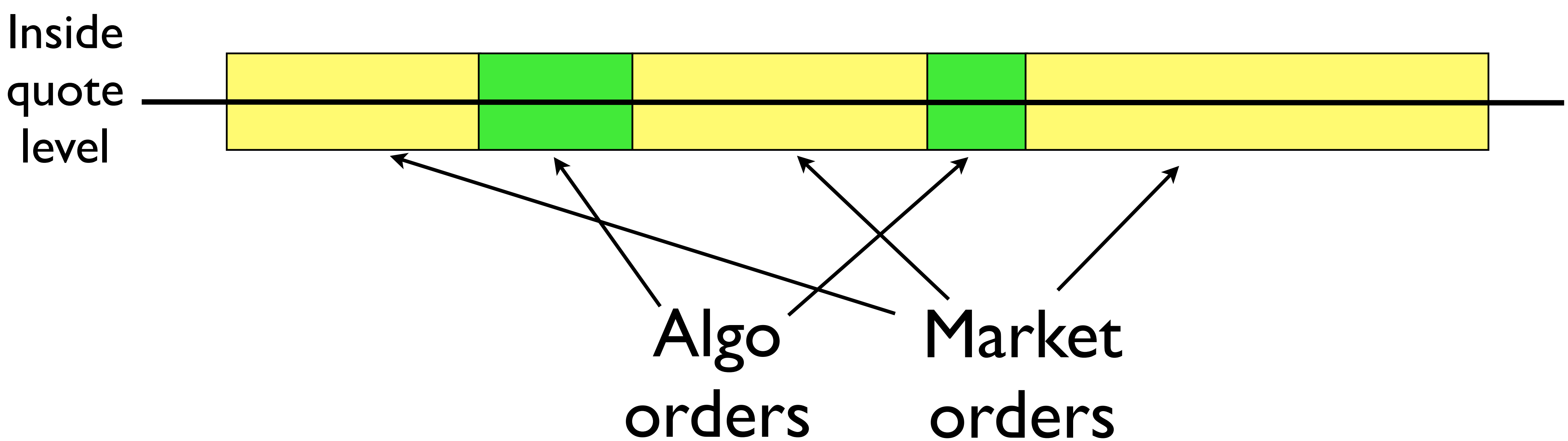
Merge real market data



Criteria for simulator

- If no algo orders, reproduce market data
- If no market data, reproduce match engine
- Challenge: combine market data with orders

- Interleave algo orders with market



Interleave algo orders and market orders, respecting time priority and implementing exchange match rules

Mkt data	Algo order	Book	
quote=100 lot		100: 100	
	40 lot bid	140: 100 | 40	
quote=150 lot		190: 100 | 40 | 50	
trade=30 lot (20%) quote=120 lot	8 lot fill	trade=38 lot 152: 80 | 32 | 40	pro rata
quote=110 lot (10 lot cancel)		142: 80 | 32 | 30	cancel from back of queue
etc			

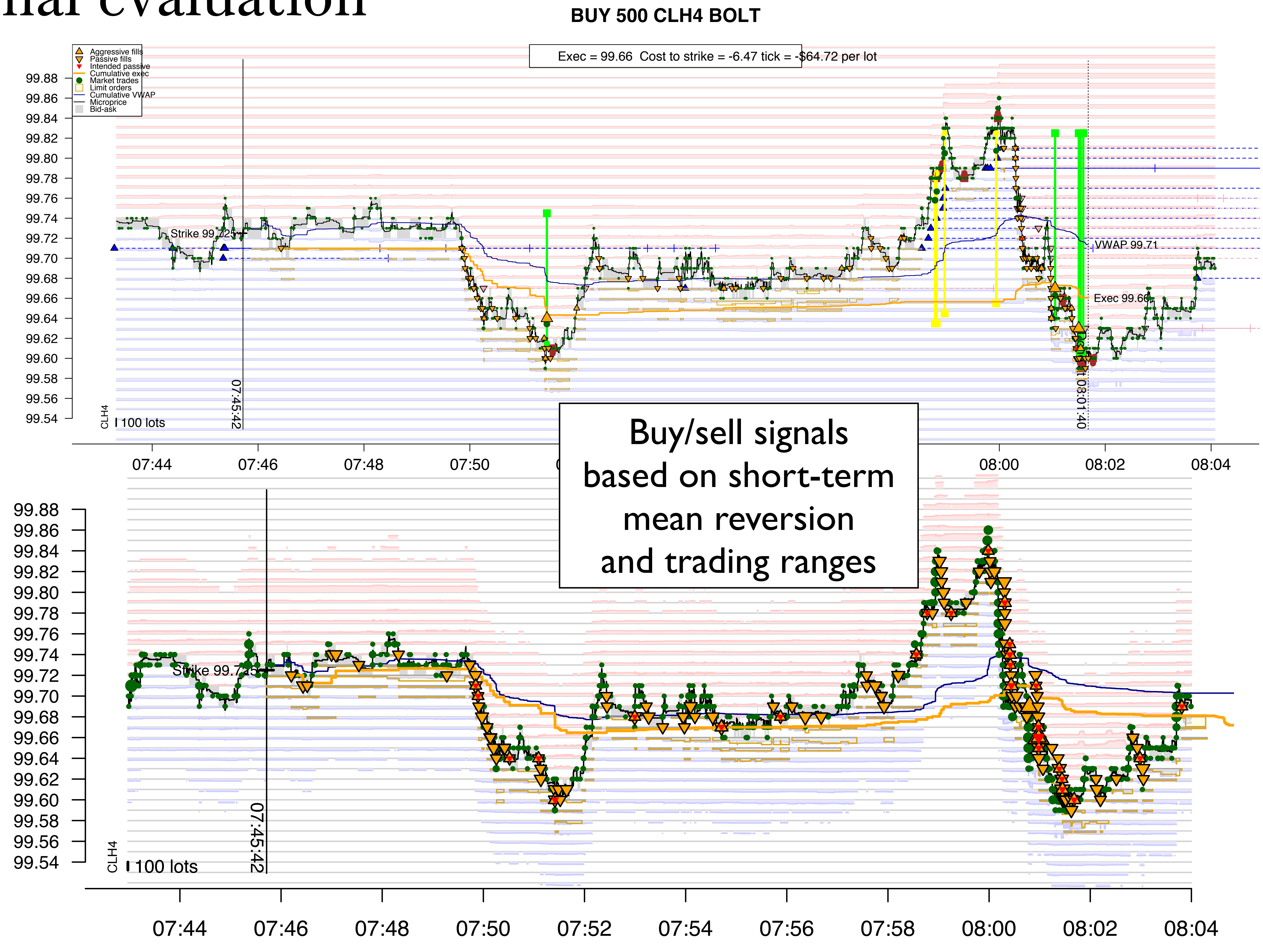
Simulator Assumptions

- Child orders always joining back of the queue
- Child orders use pessimistic queue position model, where;
  - Market Trades - reduce quantity from front of queue
  - Market Quote decreases - reduce quantity from back of queue
- Child orders receive passive fills based on matching algorithm:
- Aggressive child orders are fully executed at sweep price
- Child orders cannot establish a new price level
- If a price level is traded through, child orders at that level are filled
- Hidden liquidity (BML) is recreated from QB calculations
- Implied quotes are treated equally to direct quotes
- Static latency of 2ms on market data and 8ms on execution

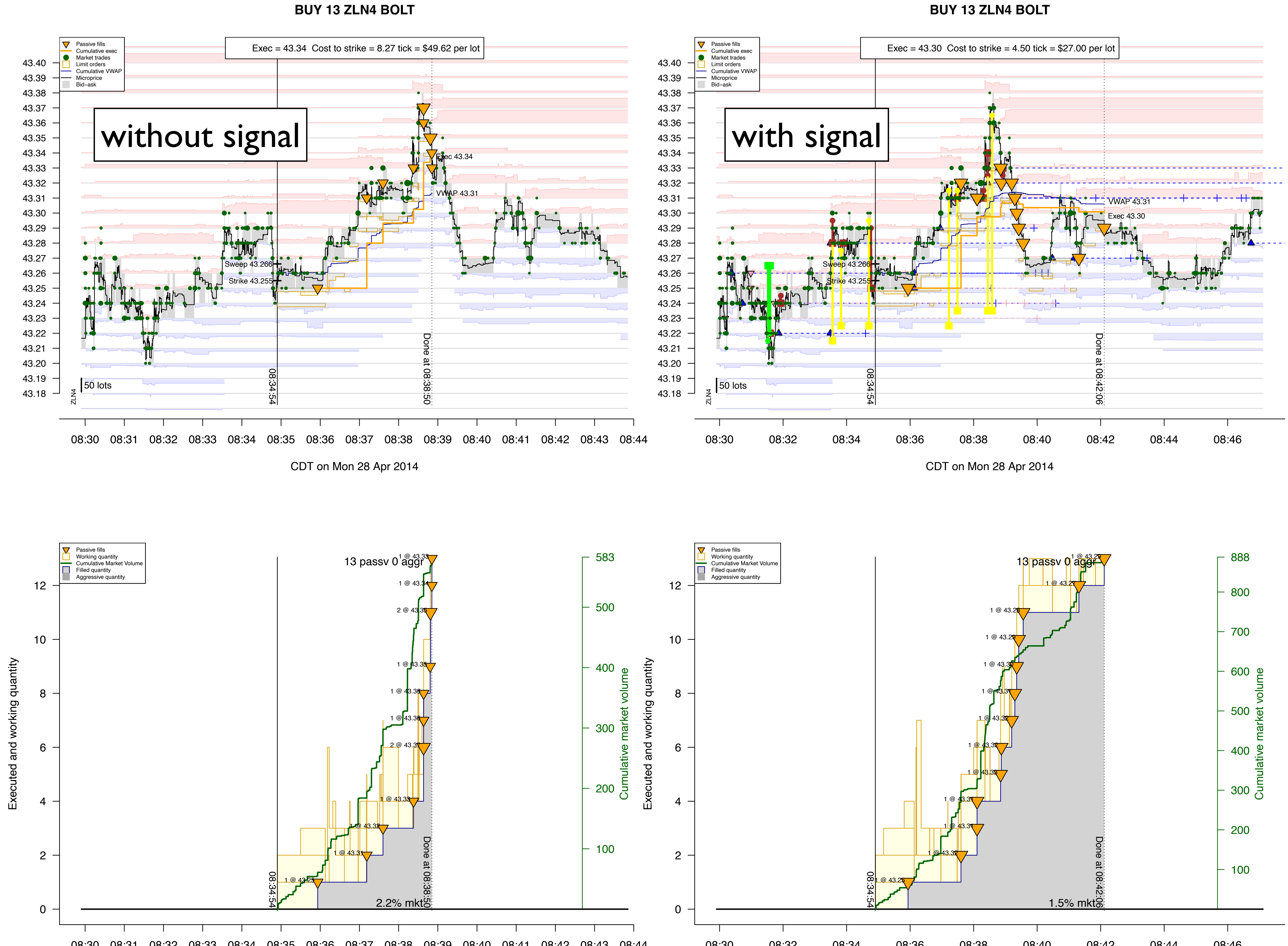
How to use simulator

- Historical
  - rerun scenarios for algo improvement
  - backtests for potential clients
- Real-time
  - clients can connect to “test-drive” algos
- Algorithm development
  - test new signals on historical orders
  - multi-market legging trades
- Real-time splitting for testing
  - compare simulator executions with real

Signal evaluation



Buy/sell signals based on short-term mean reversion and trading ranges



Produced by QB from CME and internal data

Remainder of today

1. Bid-ask imbalance as an example signal
2. Framework for evaluating signal significance

Bid-ask imbalance

- Most useful for large-tick assets

Bid and ask prices p_{B}, p_{A}

Bid and ask quote sizes q_{B}, q_{A}

$$p_{\mathrm{B}} \;<\; p_{\mathrm{MP}} \;<\; p_{\mathrm{A}}$$

$$q_{\mathrm{B}} \to 0 \qquad\qquad q_{\mathrm{A}} \to 0$$

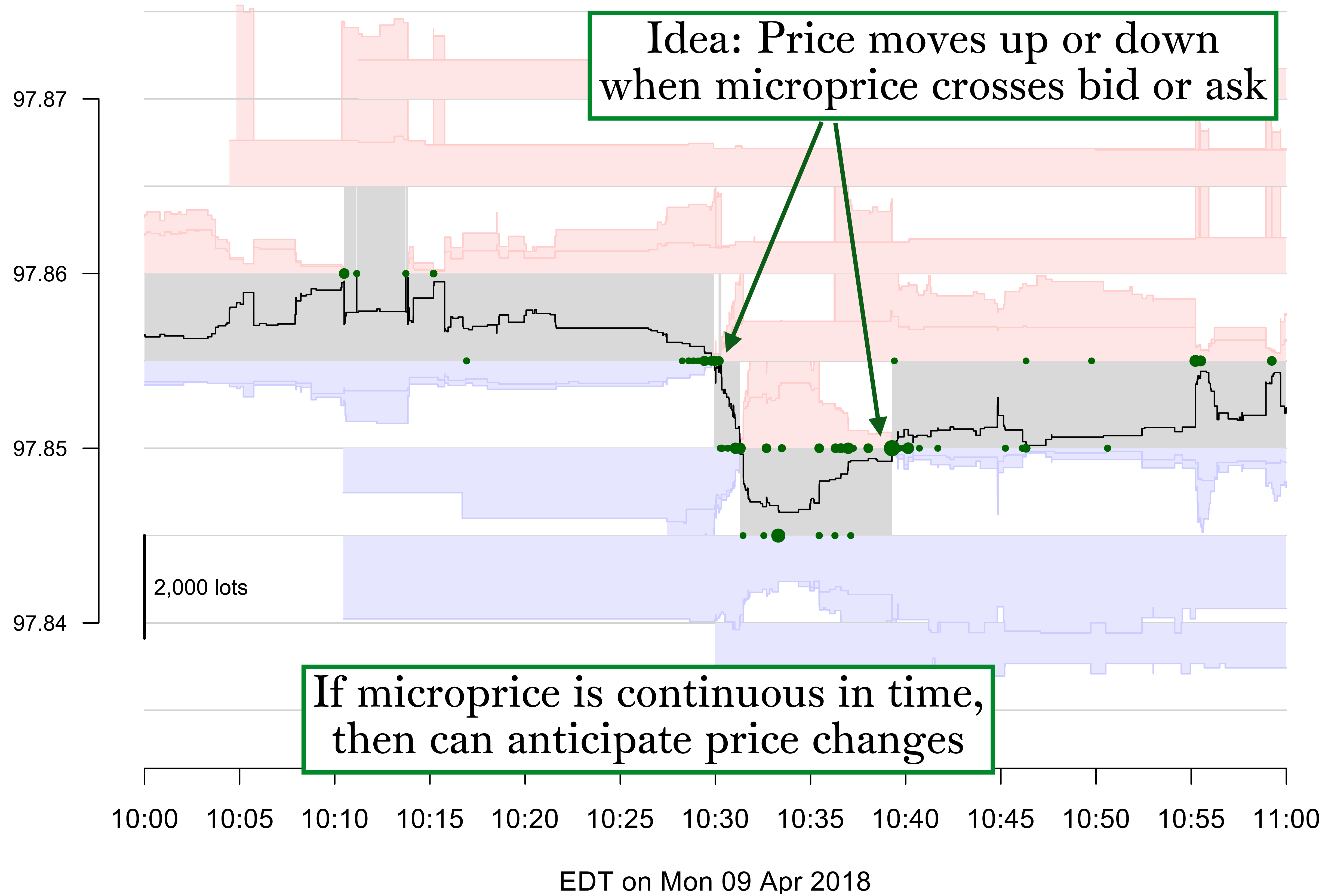
Microprice: $$p_{\mathrm{MP}} = \frac{q_{\mathrm{B}} p_{\mathrm{A}} + q_{\mathrm{A}} p_{\mathrm{B}}}{q_{\mathrm{A}} + q_{\mathrm{B}}}.$$

"Fair value"

Imbalance: $$S = 2\,\frac{p_{\mathrm{MP}} - p_{\mathrm{mid}}}{p_{\mathrm{A}} - p_{\mathrm{B}}} = \frac{q_{\mathrm{B}} - q_{\mathrm{A}}}{q_{\mathrm{B}} + q_{\mathrm{A}}} \qquad -1 < S < 1$$

Signal

BAXH8: Mar 2018 Canadian Banker's Acceptance (a short term interest rate, like Eurodollar)



Diffusion model for bid and ask sizes

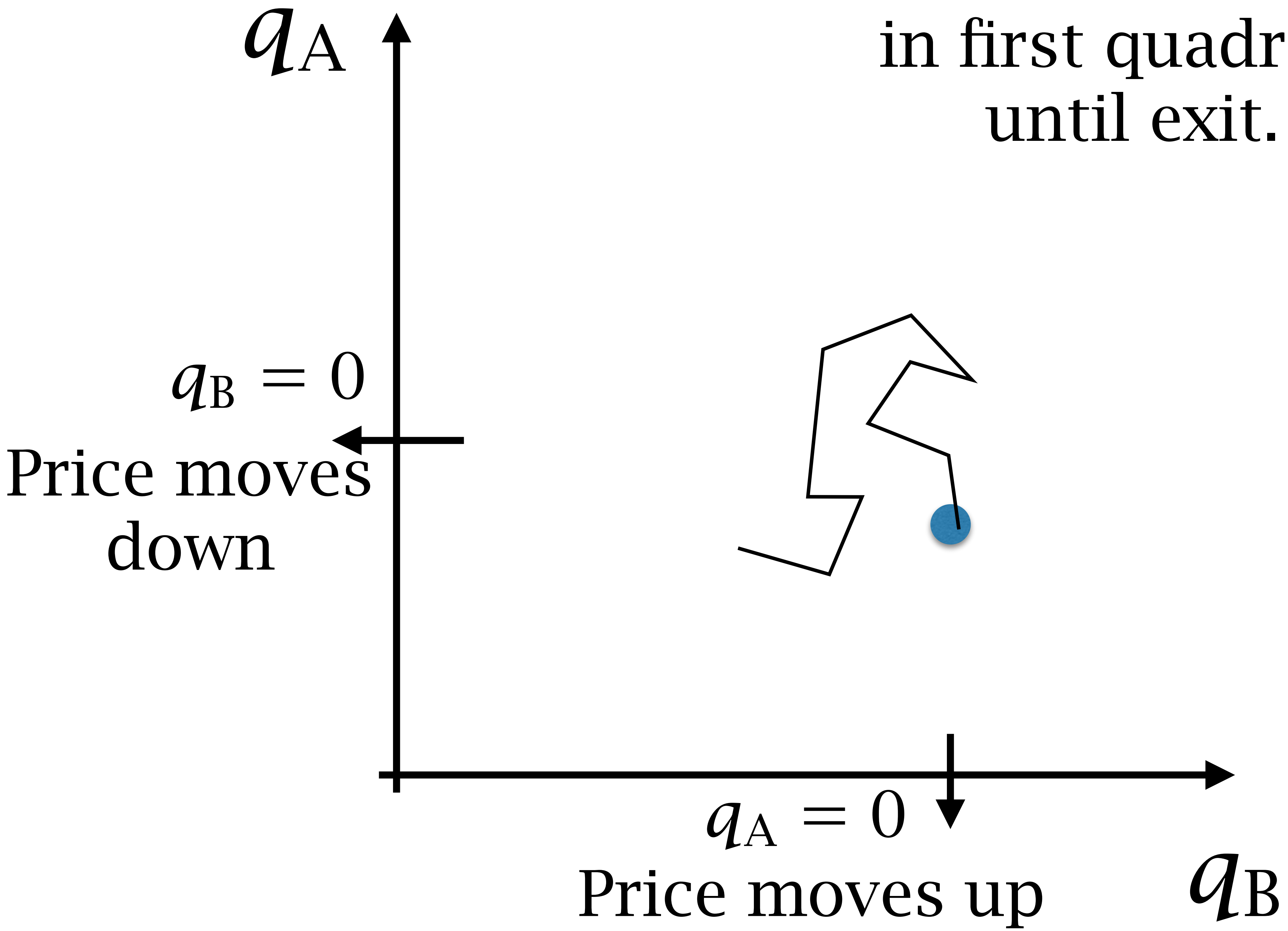
SIAM J. FINANCIAL MATH.
Vol. 4, pp. 1–25

Price Dynamics in a Markovian Limit Order Market*

Rama Cont† and Adrien de Larrard†

Abstract. We propose a simple stochastic model for the dynamics of a limit order book, in which arrivals of market orders, limit orders, and order cancellations are described in terms of a Markovian queueing system. Price dynamics are endogenous and result from the execution of market orders against outstanding limit orders. Through its analytical tractability, the model allows us to obtain analytical expressions for various quantities of interest, such as the distribution of the duration between price changes, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, *conditional* on the state of the order book. We study the diffusion limit of the price process and express the volatility of price changes in terms of parameters describing the arrival rates of buy and sell orders and cancellations. These analytical results provide some insight into the relation between order flow and price dynamics in limit order markets.

Order book state does random walk in first quadrant until exit.



Assumptions

- Limit orders, cancellations, trades arrive randomly
  independent Markov processes
- Cancellations + trades = new orders
  order book size is stationary process
- Order flow is balanced and uncorrelated
- Continuum limit
  lots of small limit orders per unit time

Calculate

- Time distribution of next move, and
- Probability of up/down move

as functions of instantaneous state (queue sizes)

Distribution of transition time

Proposition 1 (distribution of duration until next price move). *The distribution of τ conditioned on the initial state of the order book is given by*

$$\mathbb{P}[\tau > t | q_0^b = x,\ q_0^a = y] = \sqrt{\left(\frac{\mu+\theta}{\lambda}\right)^{x+y}} \quad \psi_{x,\lambda,\theta+\mu}(t) \quad \psi_{y,\lambda,\theta+\mu}(t), \tag{3}$$

where

$$\psi_{n,\lambda,\theta+\mu}(t) = \int_t^\infty \frac{n}{u} I_n(2\sqrt{\lambda(\theta+\mu)}u) e^{-u(\lambda+\theta+\mu)} du \tag{4}$$

and I_n is the modified Bessel function of the first kind.

- If the order flow is balanced, i.e., $\lambda = \mu + \theta$, then

  The duration then follows a heavy-tailed distribution with infinite first moment:

$$\mathbb{P}[\tau > t | q_0^b = x, q_0^a = y] \underset{t\to\infty}{\sim} \frac{xy}{\pi\lambda}\frac{1}{t}. \quad \blacksquare \tag{6}$$

Distribution of transition direction

Proposition 2. *For $(n,p) \in \mathbb{N}^2$, the probability $p_1^{up}(n,p)$ that the next price move is an increase, conditioned on having the n orders on the bid side and p orders on the ask side, is*

$$p_1^{up}(n,p) = \frac{1}{\pi}\int_0^{\pi}\left(2-\cos(t)-\sqrt{(2-\cos(t))^2-1}\right)^p \frac{\sin(nt)\cos(\frac{t}{2})}{\sin(\frac{t}{2})}dt. \tag{7}$$

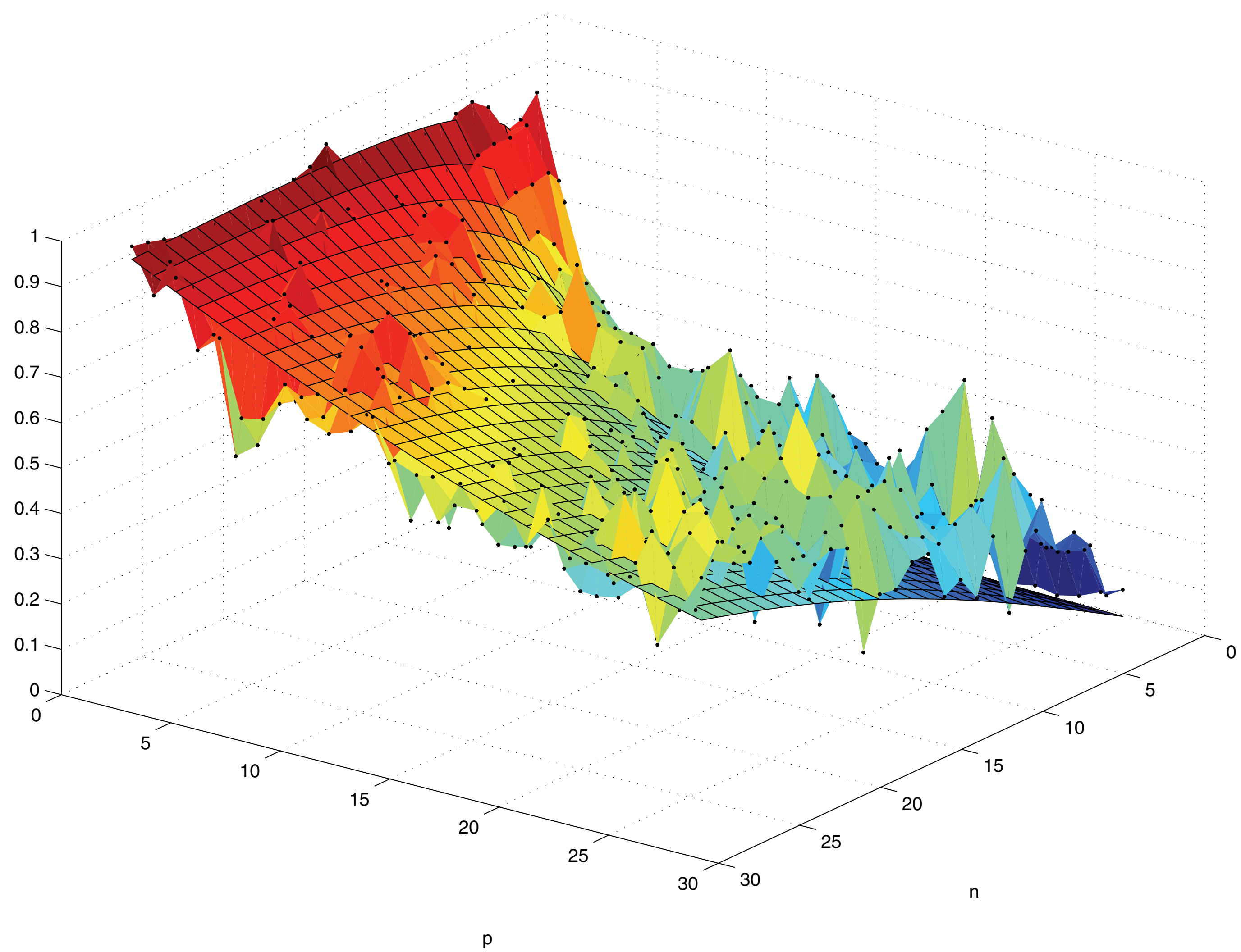
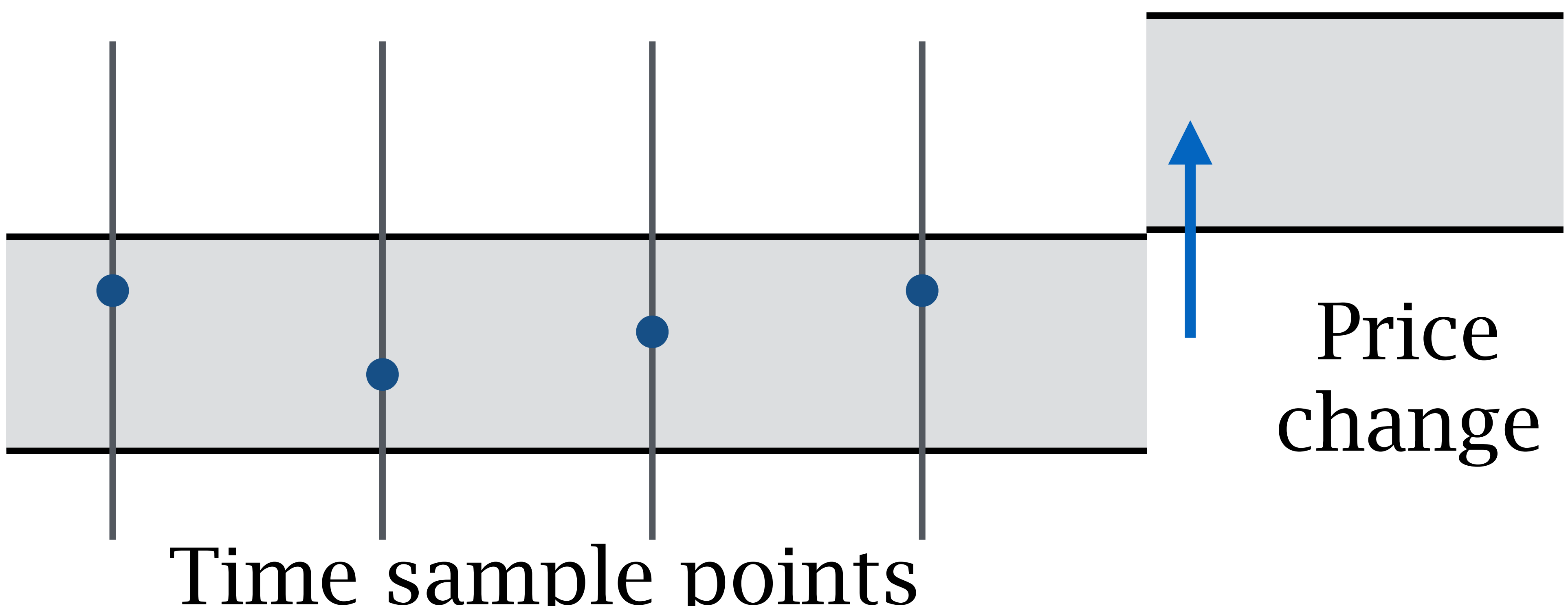
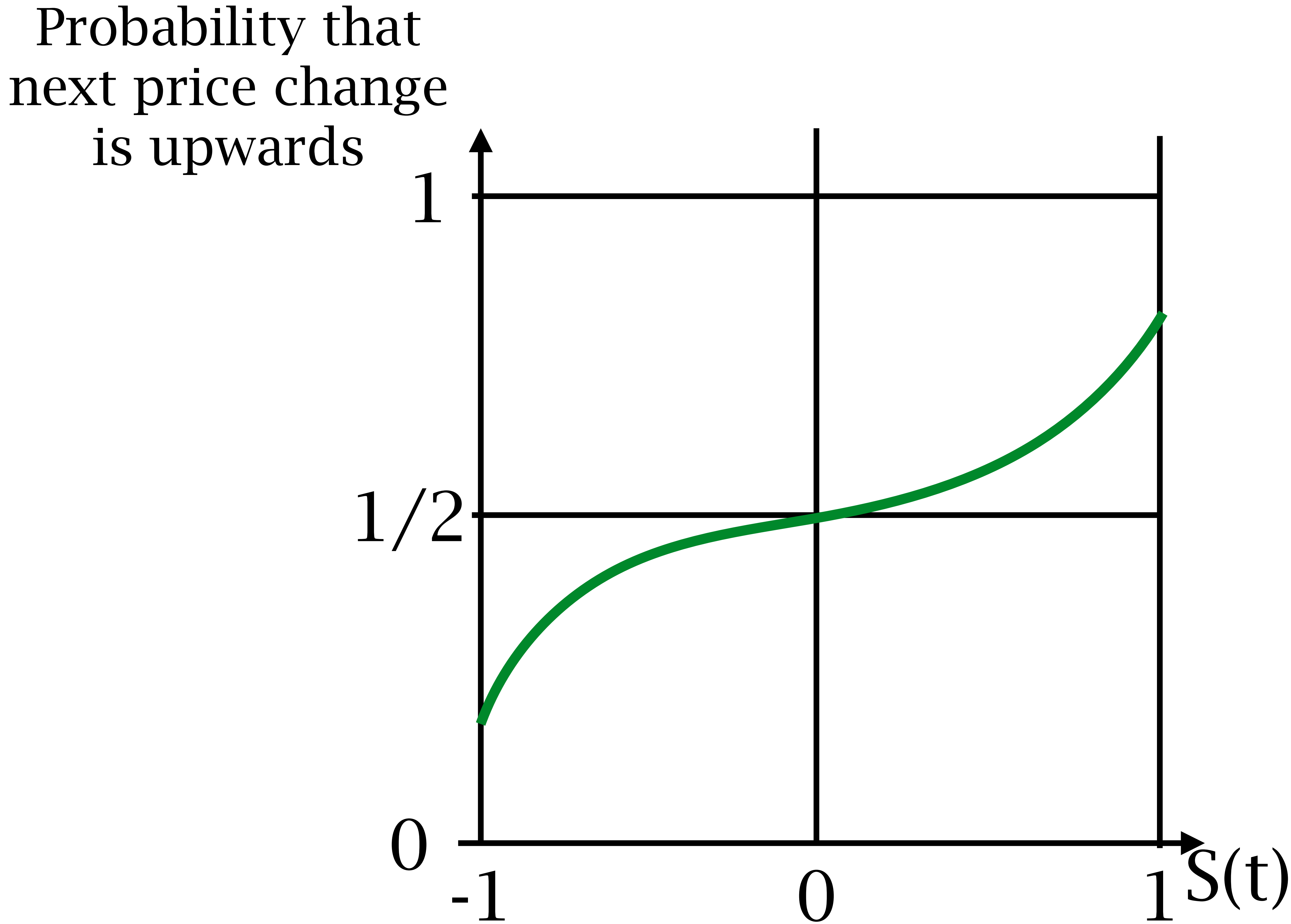


Figure 5. *Conditional probability of a price increase, as a function of the bid and ask queue sizes, compared with empirical transition frequencies for Citigroup stock price tick-by-tick data on June 26th, 2008.*

Empirical test

- Sample quotes at regular or irregular times
- Compute imbalance
- Compute relative fractions of up or down move for next price change
- Important not to condition on future (e.g. last microprice value before price change)

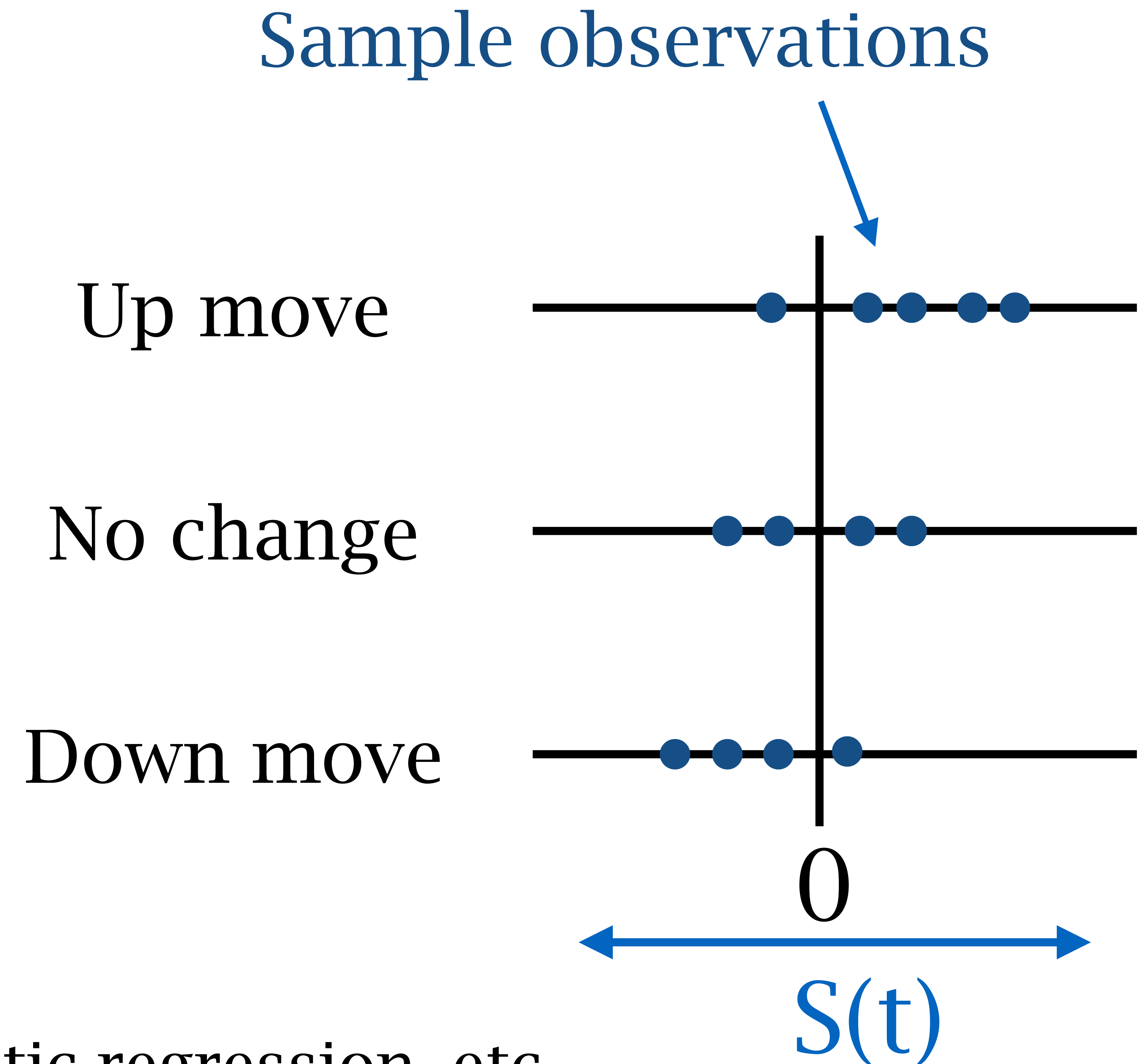




Probability that next price change is upwards
1
1/2
0
-1
0
1
S(t)

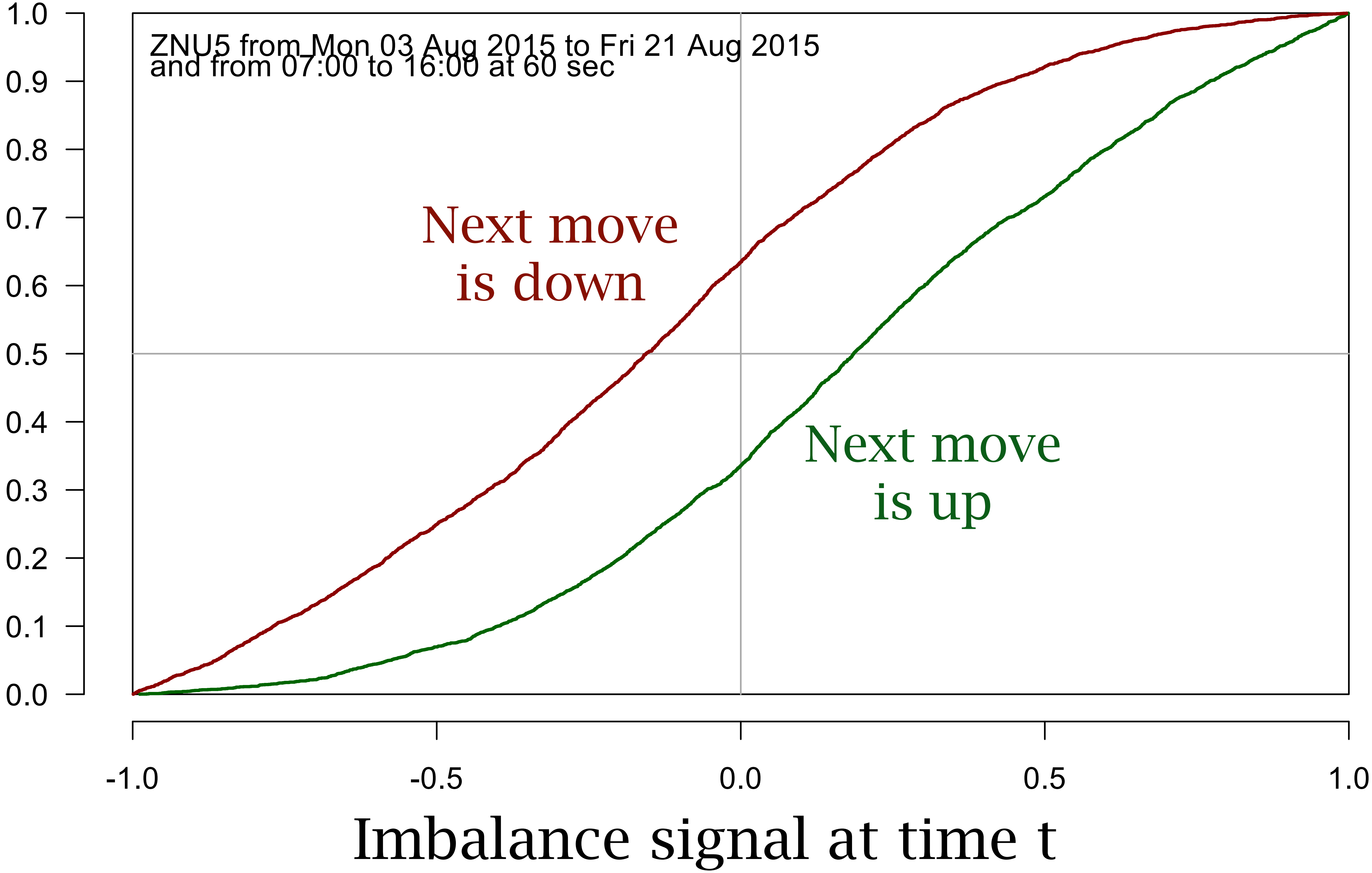
More generally

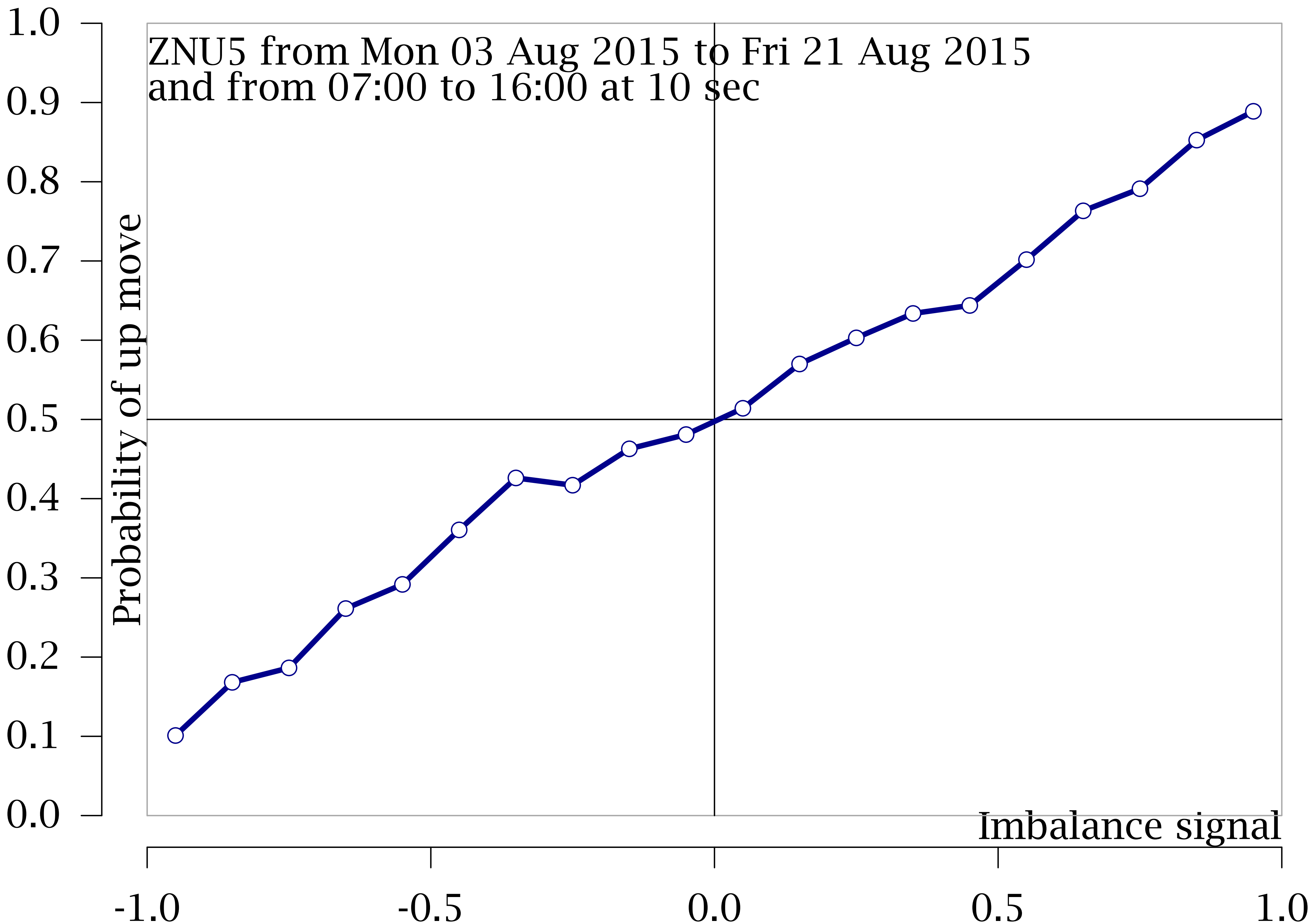
First price change within 1 minute (for example)



Can use ordinal logistic regression, etc

Cumulative distribution



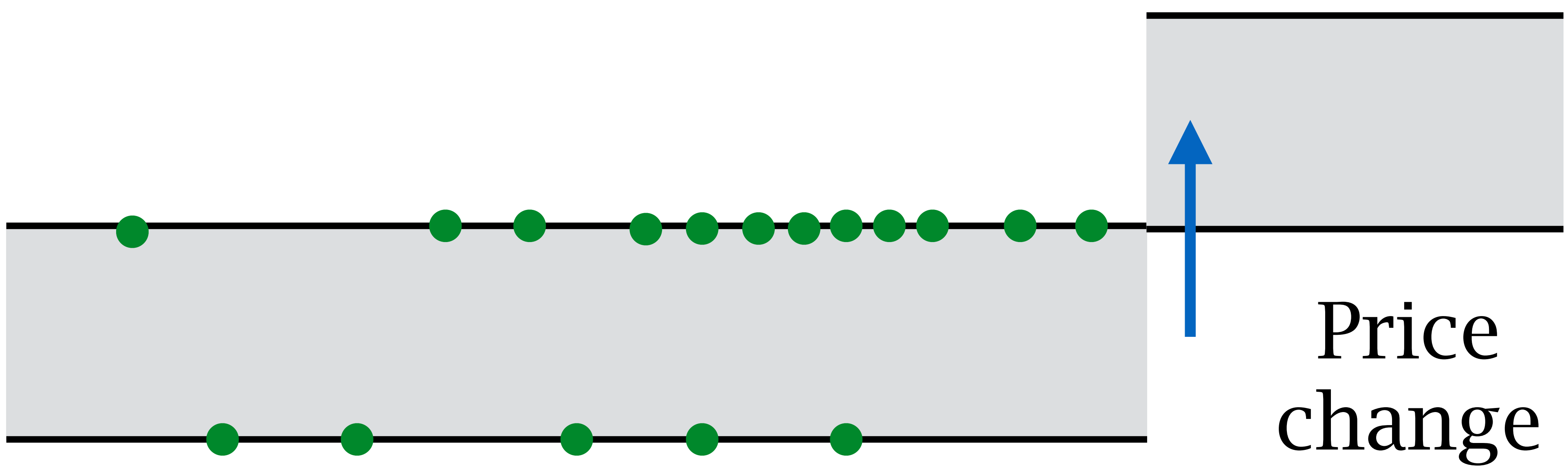

ZNU5 from Mon 03 Aug 2015 to Fri 21 Aug 2015
and from 07:00 to 16:00 at 10 sec
Probability of up move
Imbalance signal
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-1.0
-0.5
0.0
0.5
1.0

Practical questions for use of microprice

- What is the accuracy in predicting next change?
- Is one next change enough? Should it be "stable"?
- What is the distribution of waiting time?
  If very long time, not useful signal
- Where should you set your threshold for acting?
  If signal is weak, prefer to wait for passive fills
- Can be "spoofed" since popular signal
  orders at inside quotes have risk of fill
  orders deeper in the book can mislead with less risk

Related: trade imbalance

More trades at ask than at bid
suggests buy interest
suggests future upward move



Imbalance example

Development and Usage of Short Term Signals in Order Execution

Michael G Sotiropoulos

Algorithmic Trading Quantitative Research
Bank of America Merrill Lynch

Cornell Financial Engineering Seminar
New York, 10-Oct-2012

Bank of America Merrill Lynch

Signal Examples: Trade Sign Autocorrelation (III)

Calculation of the signal (Almgren 2006)

1. For each trade define its "askness" a and "bidness" b as the distance of the transaction price from the bid (res. ask) in units of spread

$$a = \min\left(\left(\frac{P - P_b}{P_a - P_b}\right)^+, 1\right); \quad b = \min\left(\left(\frac{P_a - P}{P_a - P_b}\right)^+, 1\right). \tag{21}$$

   By construction, $a + b = 1$. A trade that hits the ask side (BUY) has $a = 1, \; b = 0$.

2. At each trade time t_n compute the moving average of askness and bidness over a window of size $\tilde{\tau}_w$ as

$$A_n = \frac{1}{\tilde{\tau}_w} a_n + w_n A_{n-1}; \quad B_n = \frac{1}{\tilde{\tau}_w} b_n + w_n B_{n-1}, \tag{22}$$

   with exponentially decaying weights $w_n = e^{-\left(t_n - t_{n-1}\right)/\tilde{\tau}_w}$.

3. Normalize the moving averages by half the average trading speed

$$\bar{A}_n = \frac{2A_n}{N_{trd}/T_{day}}, \quad \bar{B}_n = \frac{2B_n}{N_{trd}/T_{day}} \tag{23}$$

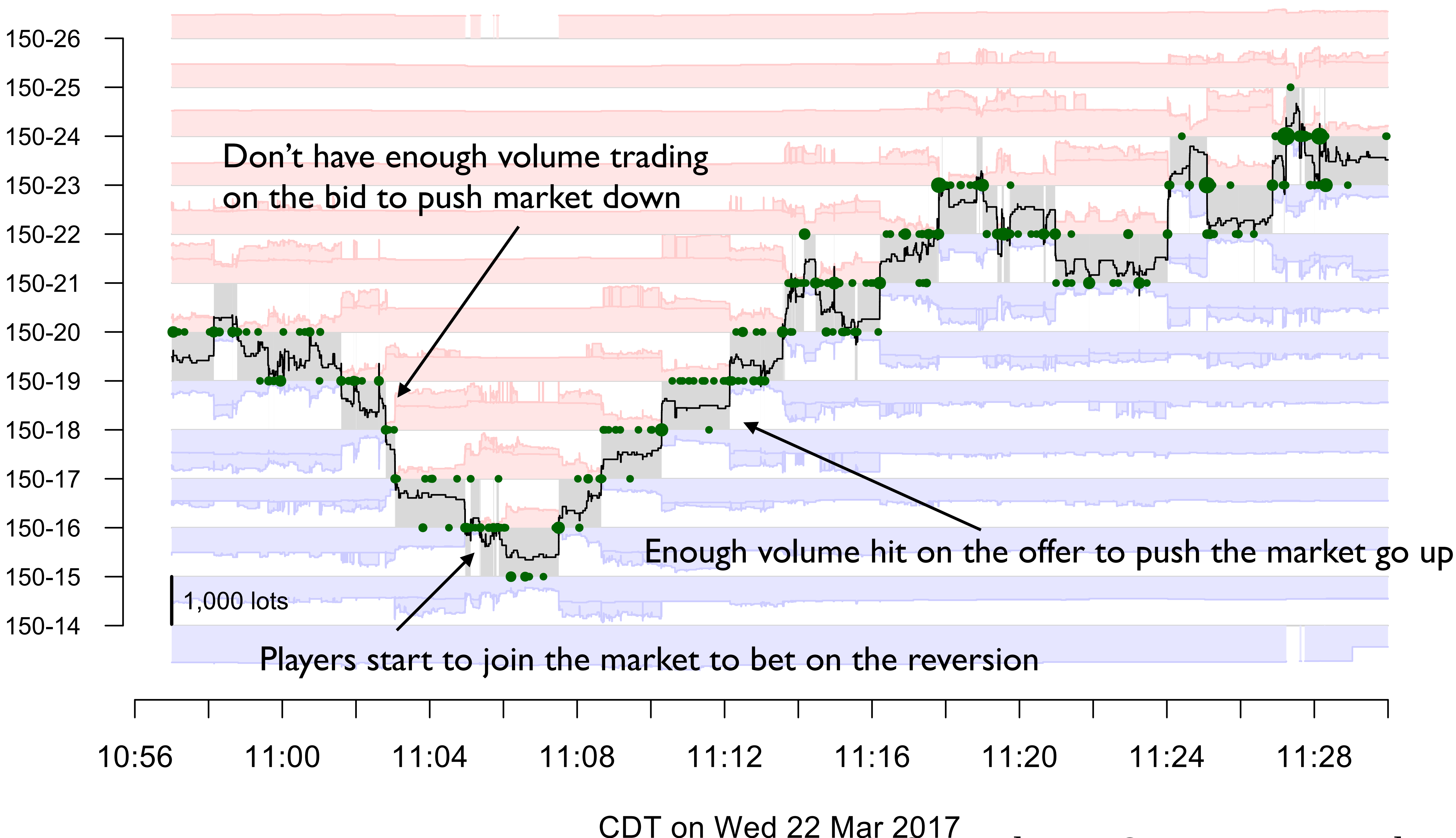
An algorithm that tries to minimize impact cost will use the signal as follows:

- For a BUY (SELL) order trade faster when $\bar{B}_n$ ($\bar{A}_n$) is higher.

Case Study from

Sean Zhang, Quantitative Brokers

1. Observe executions and market data
2. Formulate a hypothesis
3. Translate into testable signal



Sean Zhang, Quantitative Brokers

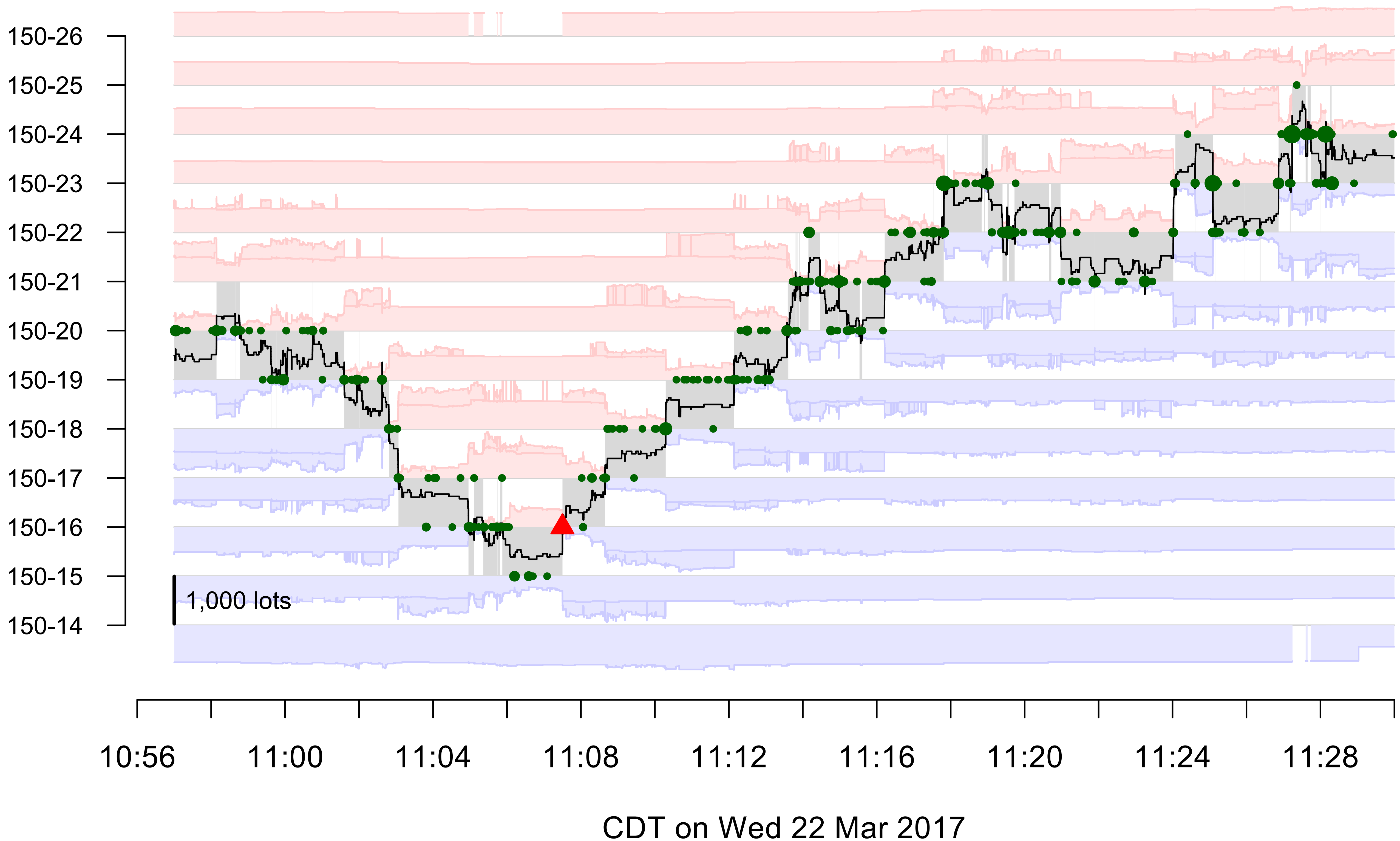
Hypothesis:
When the price drifts on light volume, then volume prints, the price will revert.

Sean Zhang, Quantitative Brokers

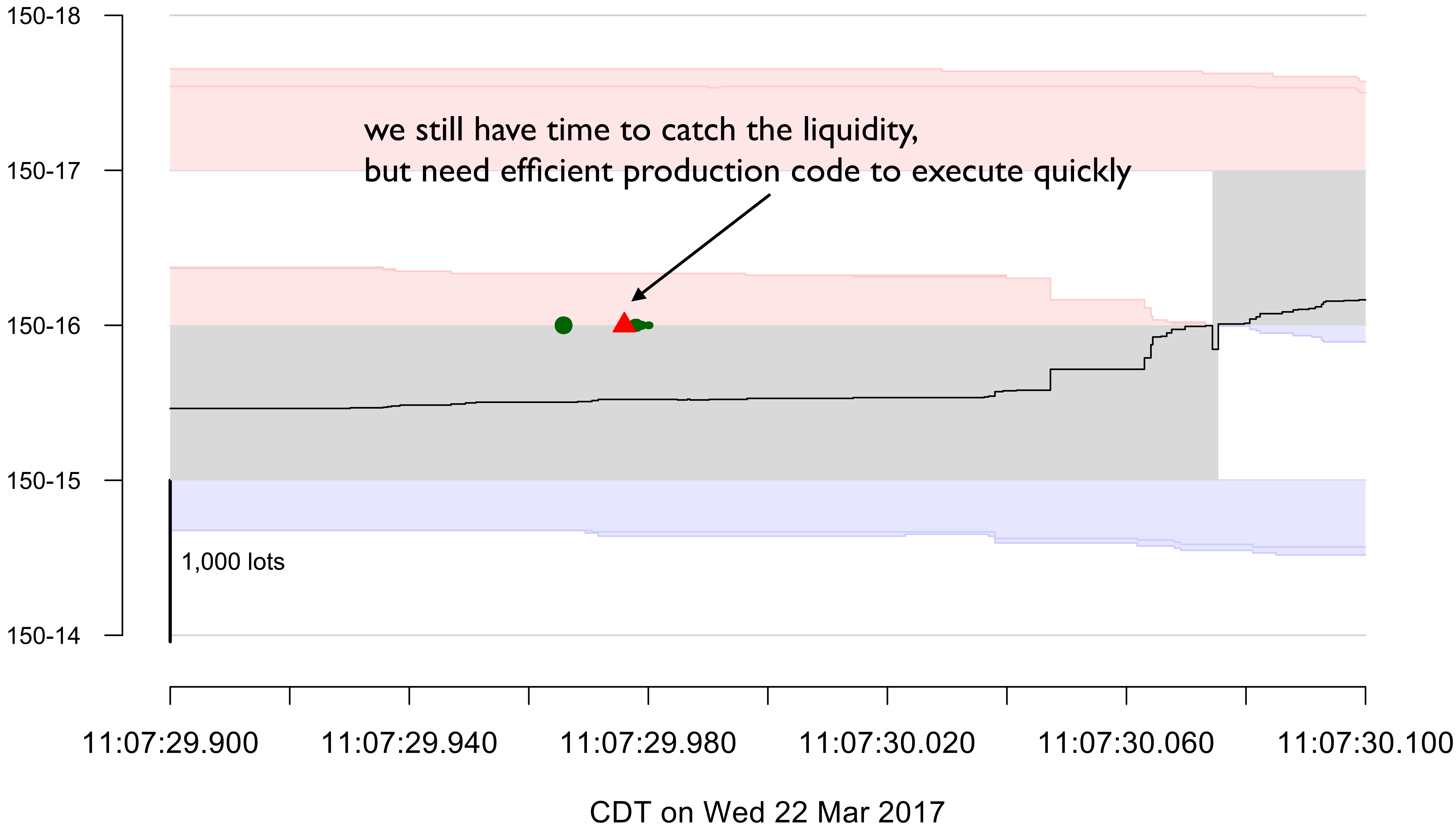
3 ways to measure volume

1. Compare with previous average traded volume at each price level
2. Compare average trade size with average quote size at each price level
3. Compare trade volume on bid and offer sides

Sean Zhang, Quantitative Brokers



Sean Zhang, Quantitative Brokers



Sean Zhang, Quantitative Brokers