PRINCETON UNIVERSITY

ORF 474: High Frequency Trading
Spring 2020
Robert Almgren

Lecture 7b

March 25, 2020

Today

- Q and R
  - functions and tables
  - forming queries from R
- Time and price sampling
- Tick size (continue from last week)
- High frequency volatility (begin)

Q and R

- Tables in Q
- Functions in Q
- Forming queries from R to Q (strings)

Goals with data analysis

- Properties of underlying "true" price process
  - volatility and correlation
- Structure of discrete market system (microstructure)
  - tick size: is it too large or too small
  - queue dynamics

- At long time scales, details do not matter
- Take times close together to either
  - get more data for large-scale properties
    - for example, minute-by-minute volatility
  - get information about microstructure

Data is discrete

- Time is discrete, not necessarily uniformly spaced
  - market events: trade or quote updates
  - or sample on discrete grid: 1 sec, 1 min, daily, etc
- Prices are discrete, on a fixed grid
  - trade prices are integer multiples of tick size
  - quote midpoints have half the tick size
  - (microprice is not on a discrete grid)

- Extract data

$$\text{Times} \quad t_1, \ldots, t_n$$

$$\text{Prices} \quad p_1, \ldots, p_n$$

(examples soon)

- Maybe other data such as

    trade sizes

    quote sizes and imbalance

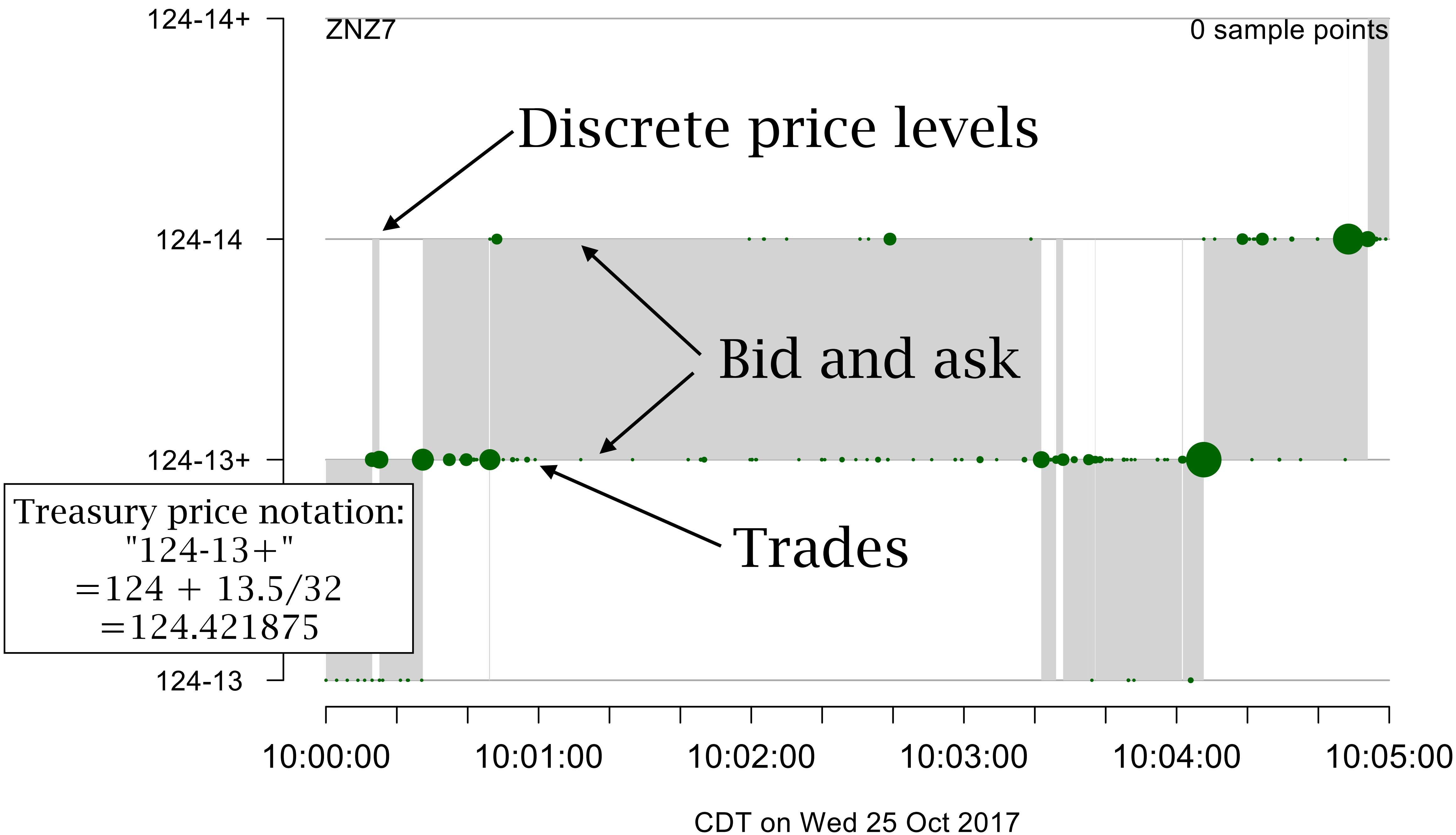
- Maybe richer data structures such as

    quote sizes at all levels in order book

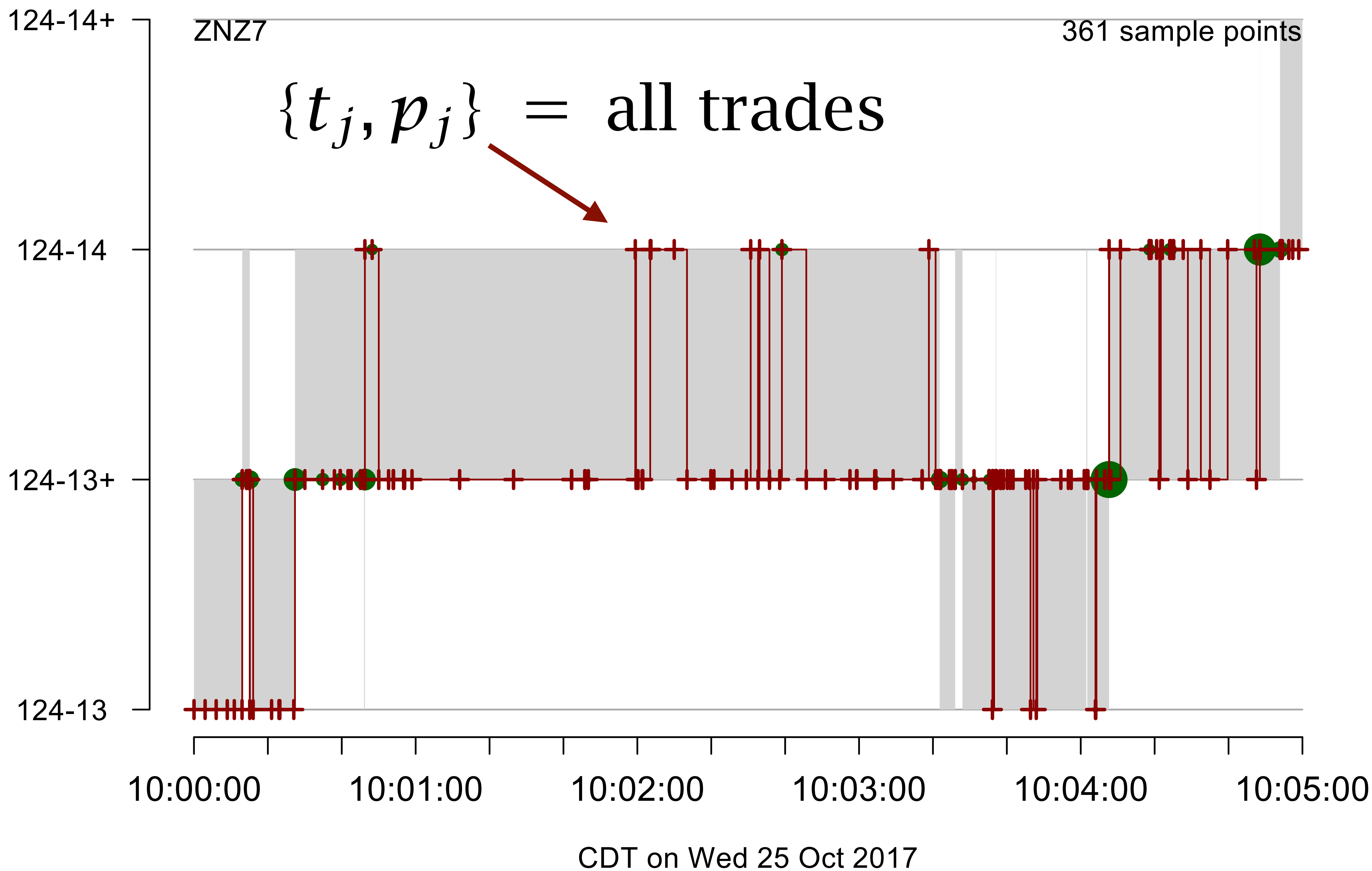
    individual orders resting in book at each price

Example for data points

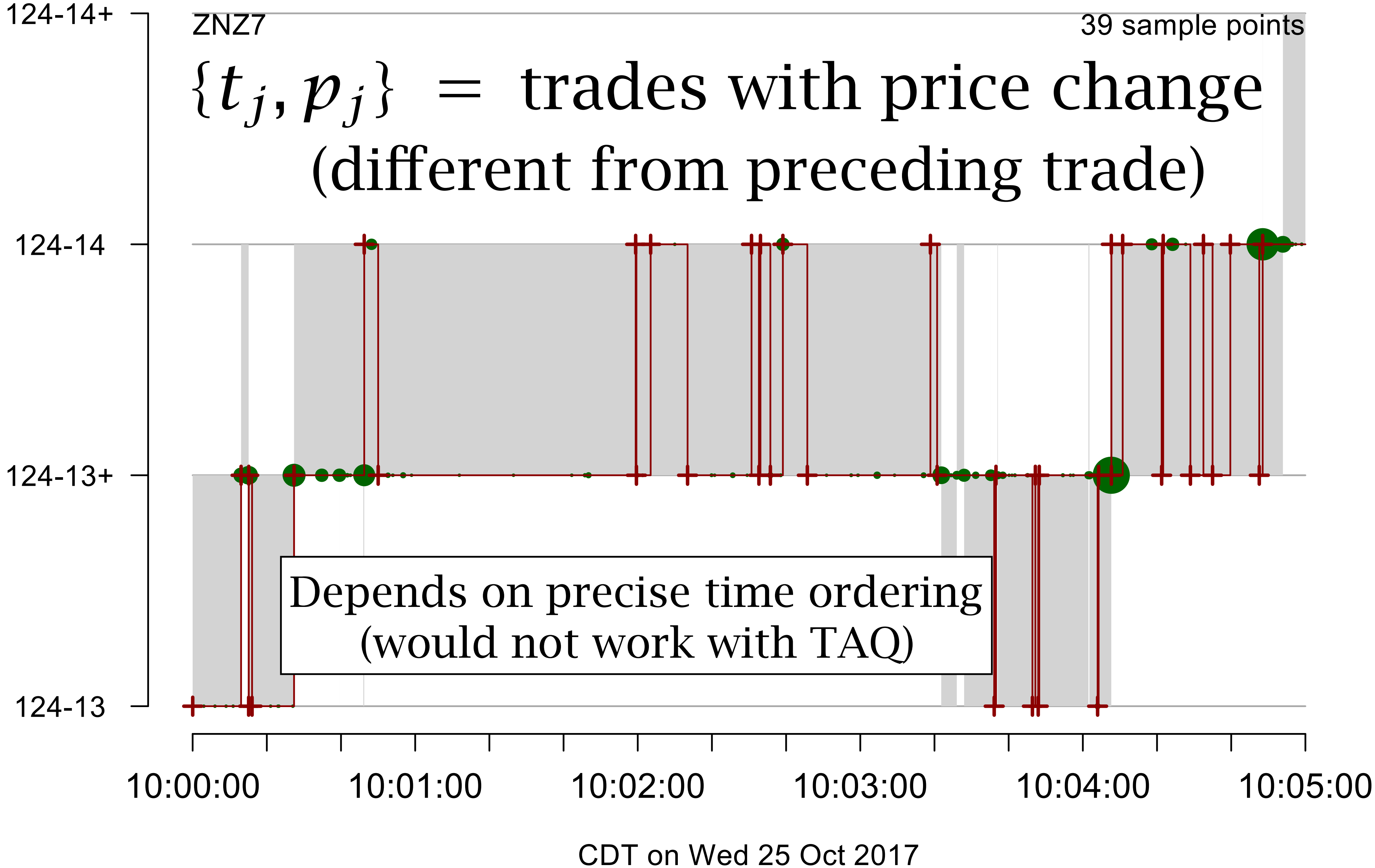
ZNZ7: 10-year Treasury futures



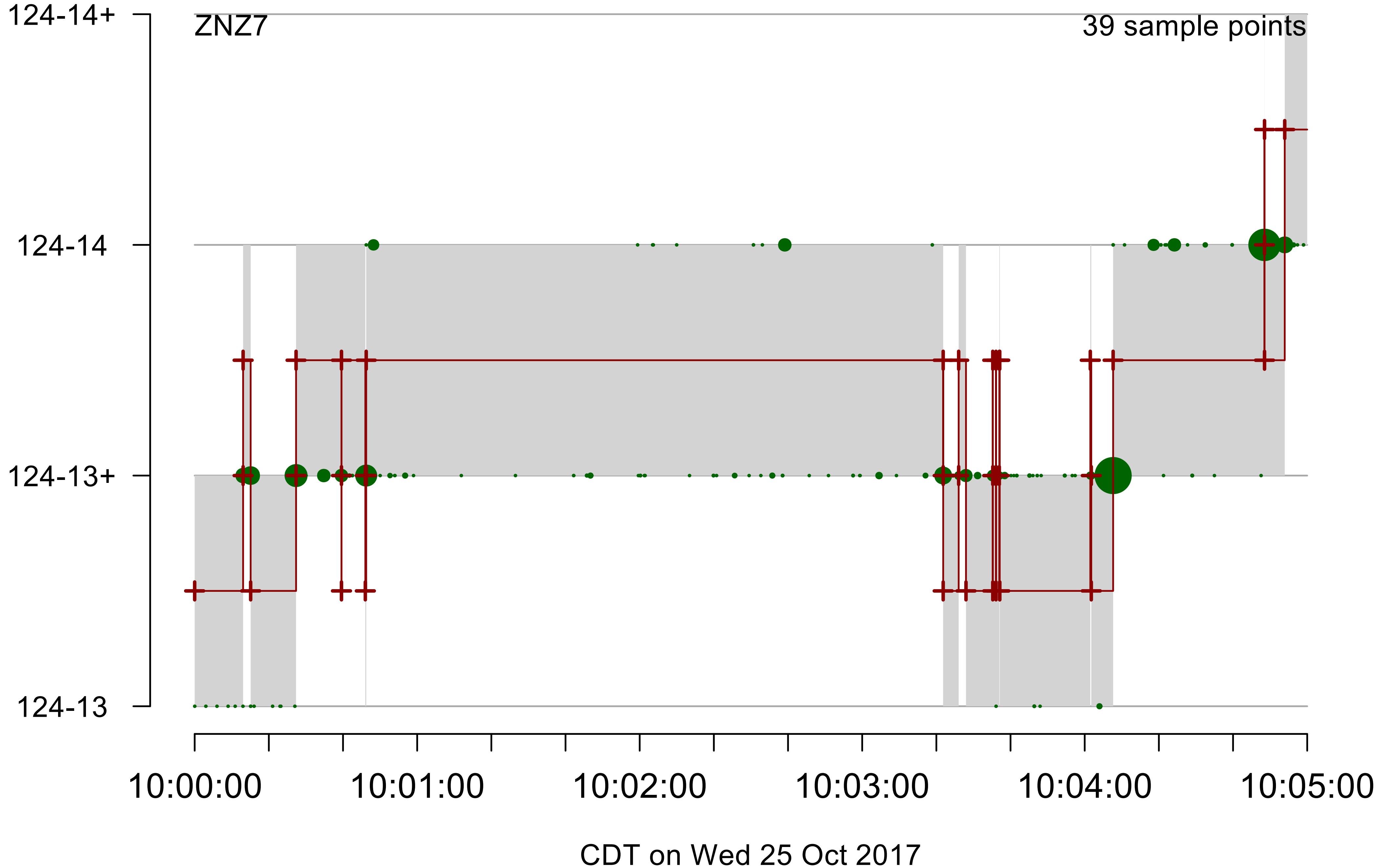
Sample every trade



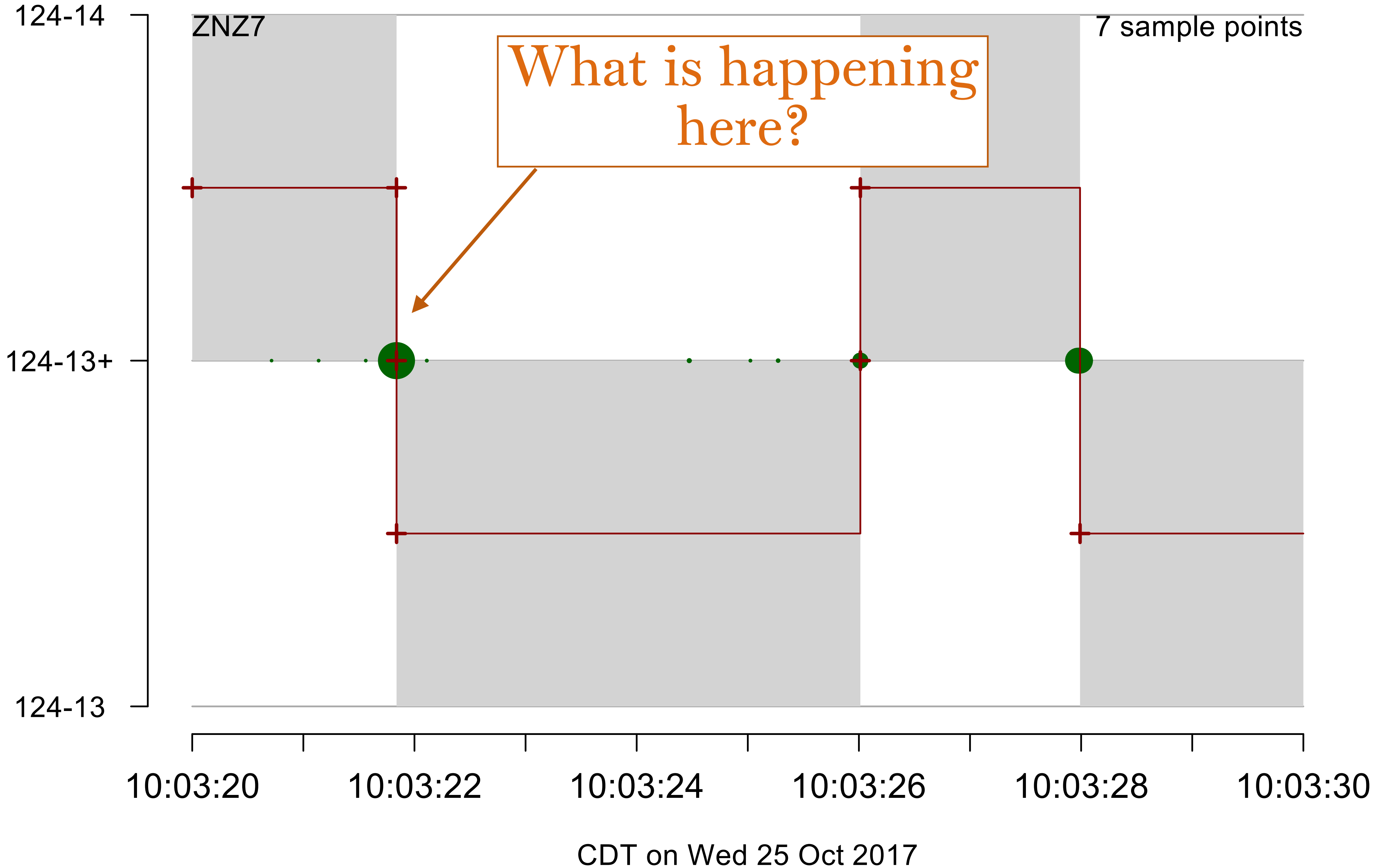
Sample at trades with new prices



Quote midpoints, when quote price changes



Detail of oscillations



Price formatting function
(defined inside overall function
so we have private variables)

```
q)h"{ pf:{(floor x)+0.01*32*x-floor x};
      update pf each bid,pf each ask,pmid:pf each 0.5*bid+ask,sprd:pf each ask-bid,
          prc: pf each prc from
      select from tqmergeT[2017.10.25;`ZNZ7;10:03:20.09;20:03:21]
           where (differ bid)|(differ ask) }[]"
date       sym  seq     time                 bsiz bid     siz prc     aggr ask     asiz pmid     sprd
-----------------------------------------------------------------------------------------------------
2017.10.25 ZNZ7 5457331 0D10:03:20.099938139 968  124.135                  124.14  2351 124.1375 0.005
2017.10.25 ZNZ7 5457369 0D10:03:20.713938877              1   124.135 S
2017.10.25 ZNZ7 5457371 0D10:03:20.713938877 967  124.135                  124.14  2351 124.1375 0.005
2017.10.25 ZNZ7 5457453 0D10:03:21.137467579              2   124.135 S
2017.10.25 ZNZ7 5457455 0D10:03:21.137467579 936  124.135                  124.14  2465 124.1375 0.005
2017.10.25 ZNZ7 5457464 0D10:03:21.561565005              2   124.135 S
2017.10.25 ZNZ7 5457466 0D10:03:21.561565005 912  124.135                  124.14  2445 124.1375 0.005
2017.10.25 ZNZ7 5457474 0D10:03:21.833148807              92  124.135 S
2017.10.25 ZNZ7 5457476 0D10:03:21.833148807 814  124.135                  124.14  2447 124.1375 0.005
2017.10.25 ZNZ7 5457507 0D10:03:21.837042131              57  124.135 S
2017.10.25 ZNZ7 5457509 0D10:03:21.837042131 486  124.135                  124.14  2484 124.1375 0.005
2017.10.25 ZNZ7 5457510 0D10:03:21.837250337              27  124.135 S
2017.10.25 ZNZ7 5457512 0D10:03:21.837250337 460  124.135                  124.14  2484 124.1375 0.005
2017.10.25 ZNZ7 5457513 0D10:03:21.837324641              9   124.135 S
2017.10.25 ZNZ7 5457515 0D10:03:21.837324641 451  124.135                  124.14  2484 124.1375 0.005
2017.10.25 ZNZ7 5457517 0D10:03:21.837346337              81  124.135 S
2017.10.25 ZNZ7 5457519 0D10:03:21.837346337 370  124.135                  124.14  2484 124.1375 0.005
2017.10.25 ZNZ7 5457521 0D10:03:21.837366619              10  124.135 S
2017.10.25 ZNZ7 5457523 0D10:03:21.837366619 359  124.135                  124.14  2484 124.1375 0.005
2017.10.25 ZNZ7 5457532 0D10:03:21.838147089              360 124.135 S
2017.10.25 ZNZ7 5457534 0D10:03:21.838147089 3251 124.13                   124.14  2831 124.135  0.01
2017.10.25 ZNZ7 5457540 0D10:03:21.838337875 1    124.135                  124.14  2841 124.1375 0.005
2017.10.25 ZNZ7 5457548 0D10:03:21.838891893              1   124.135 S
2017.10.25 ZNZ7 5457552 0D10:03:21.838891893 3241 124.13                   124.135 241  124.1325 0.005
2017.10.25 ZNZ7 5457684 0D10:03:21.839758433              4   124.135 B
2017.10.25 ZNZ7 5457686 0D10:03:21.839758433 2918 124.13                   124.135 784  124.1325 0.005
...
```

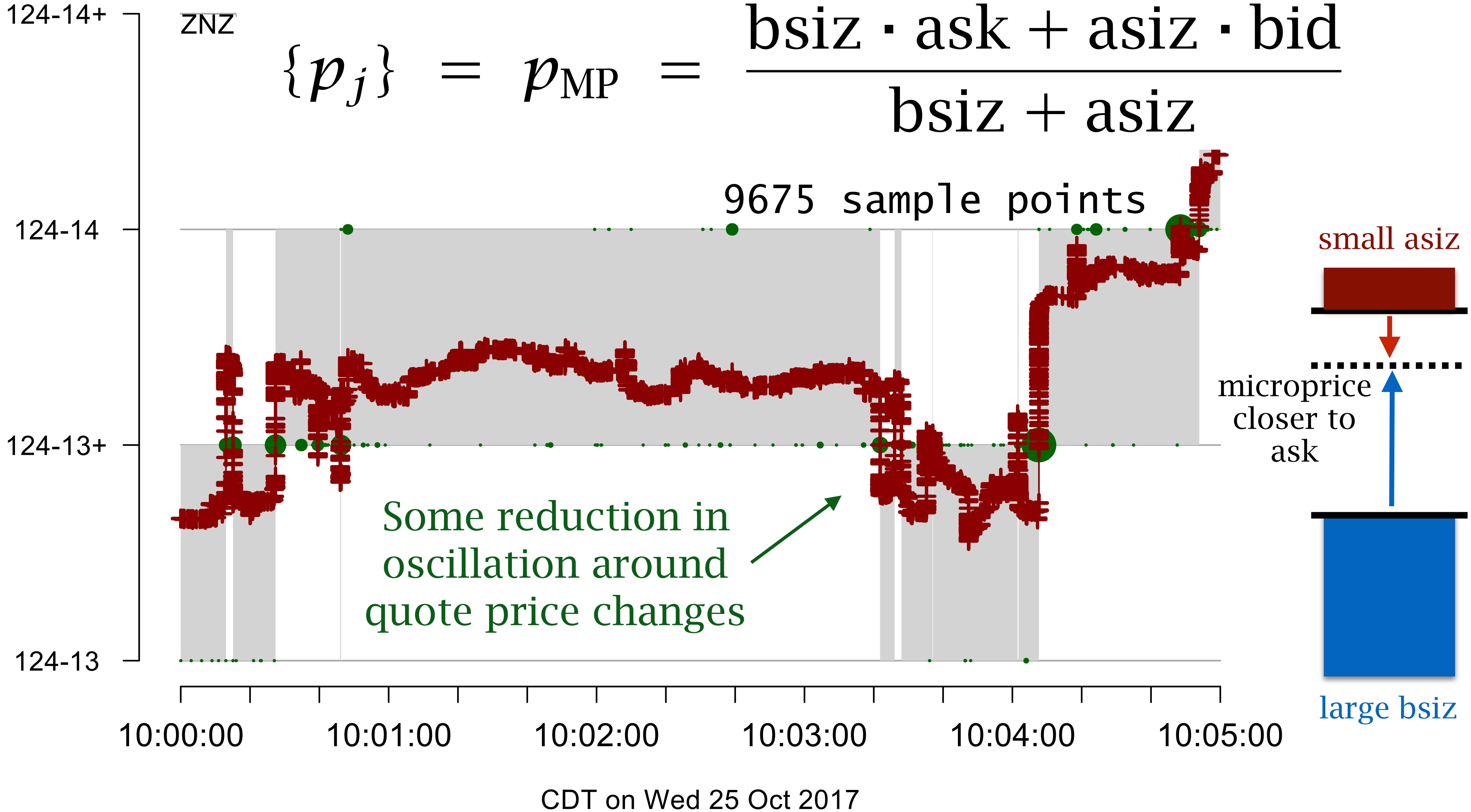
3 quotes, 2 trades
within same millisecond

"124.135" = 124 + 13.5/32

Microprice sampling

$$\{t_j\} = \text{Every quote update}$$

$$\{p_j\} = p_{\text{MP}} = \frac{\text{bsiz} \cdot \text{ask} + \text{asiz} \cdot \text{bid}}{\text{bsiz} + \text{asiz}}$$

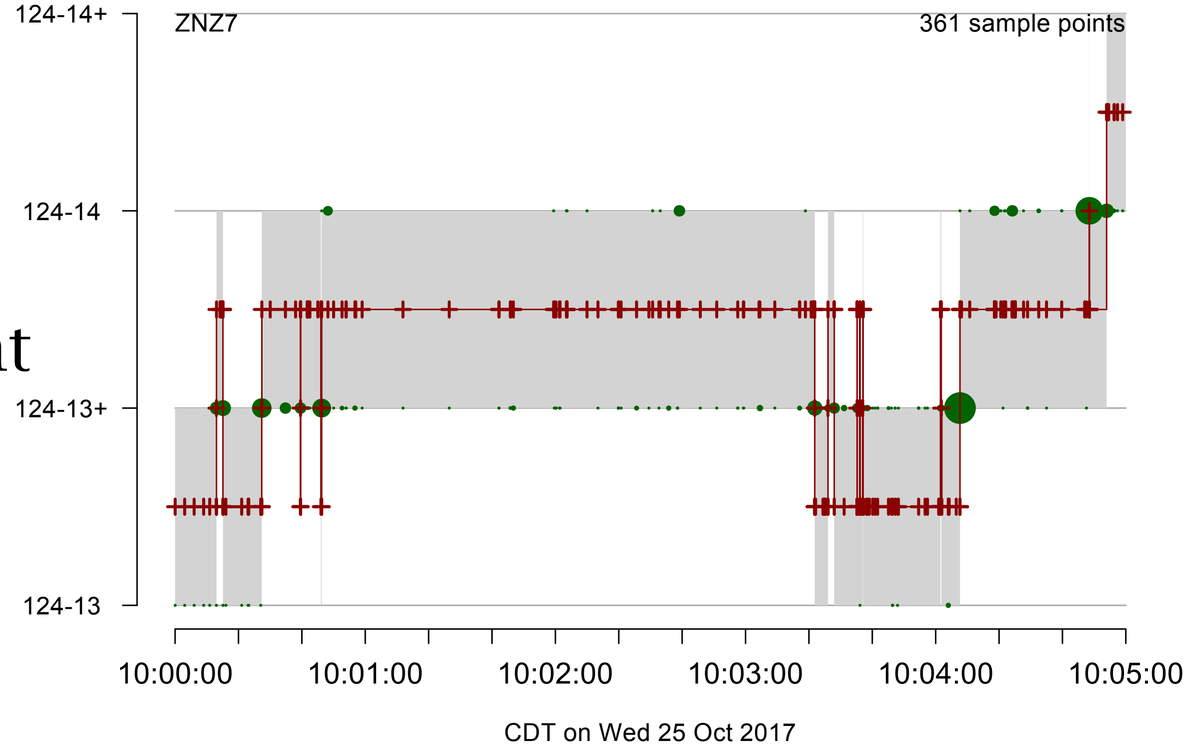


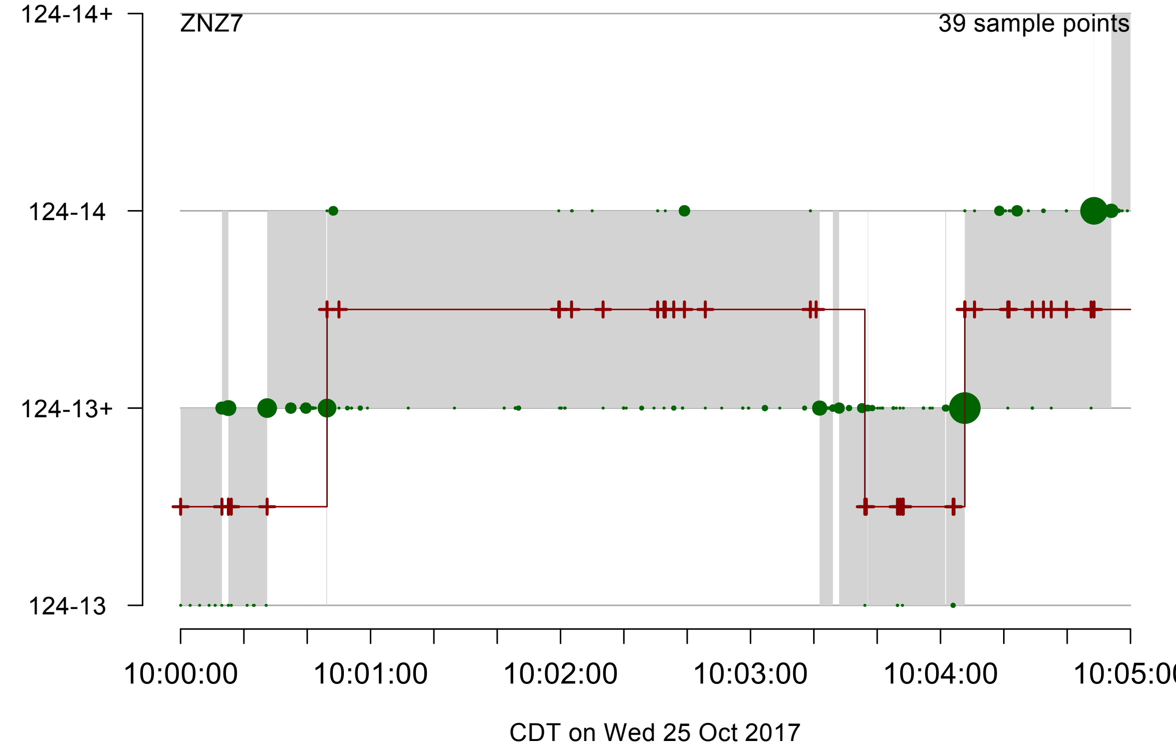
Sampling quotes at trade times

All trades

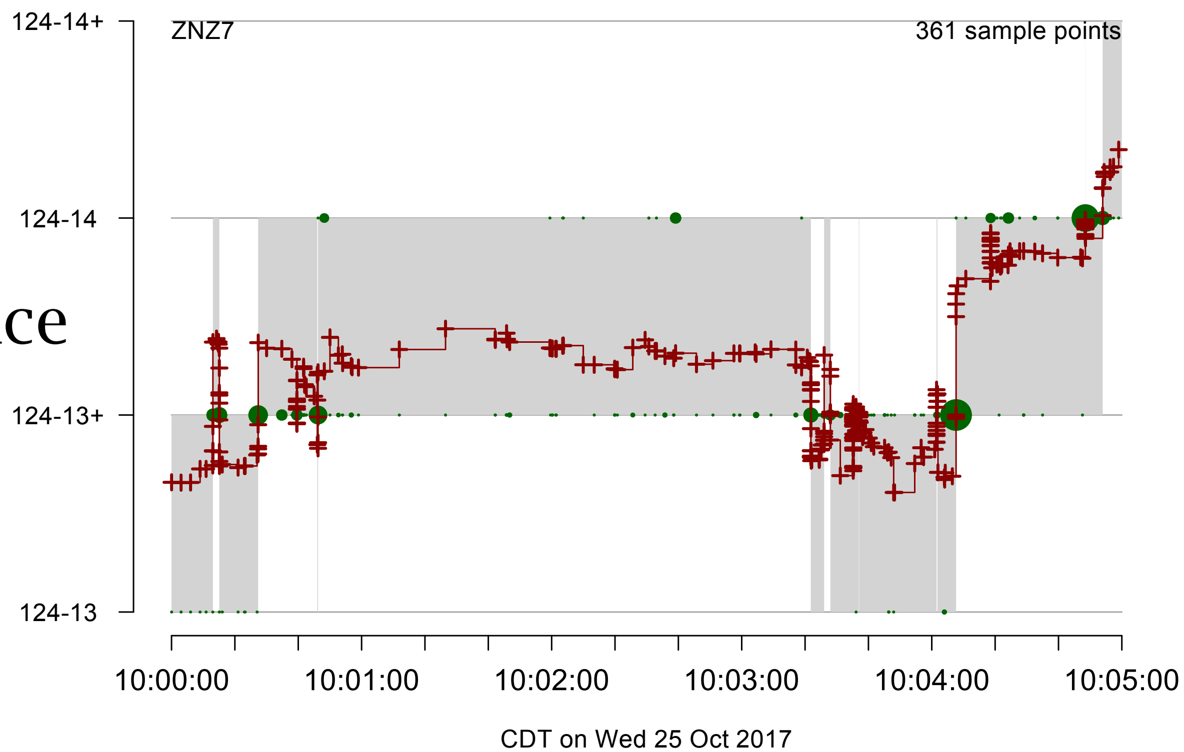
Trades with price change

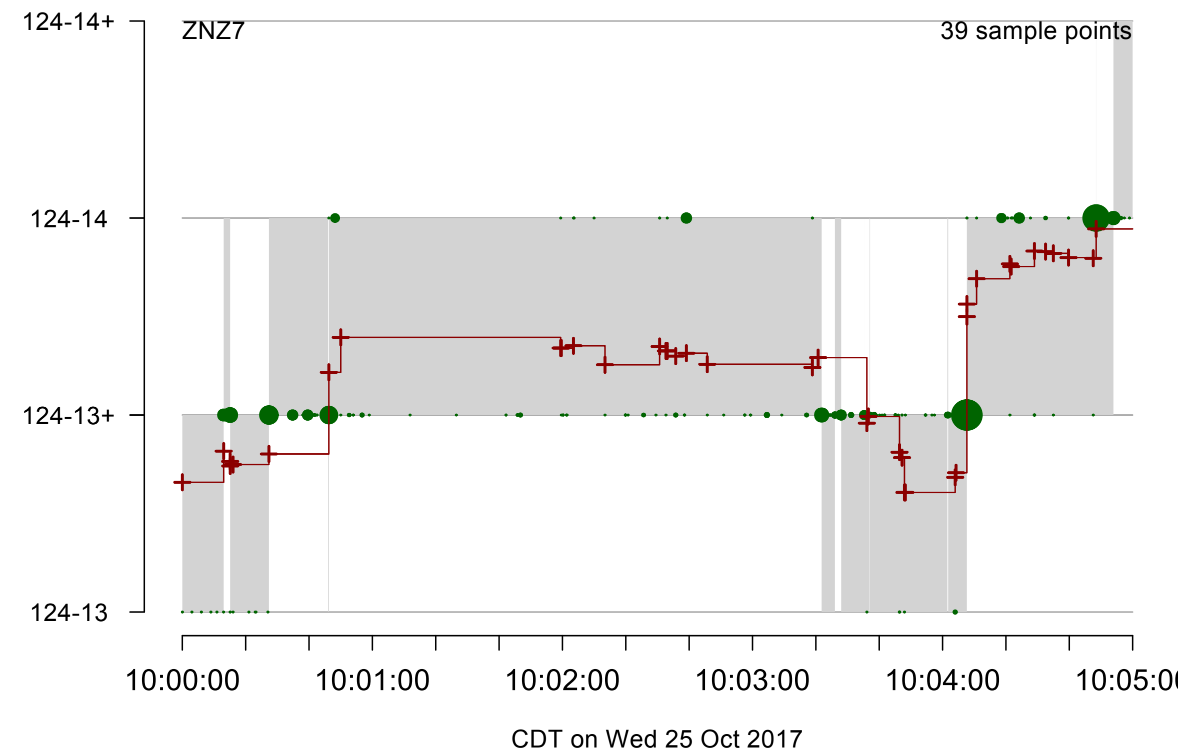
Quote midpoint





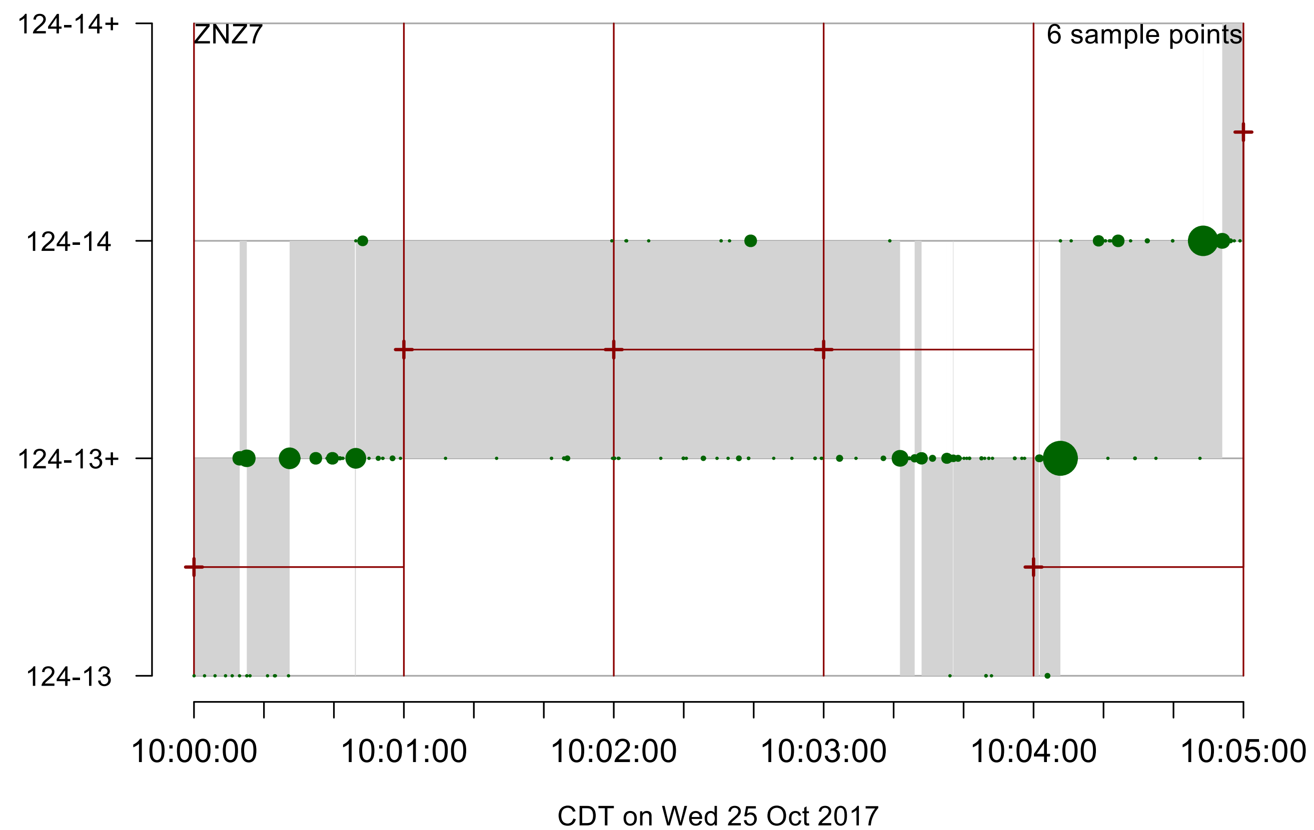
Microprice



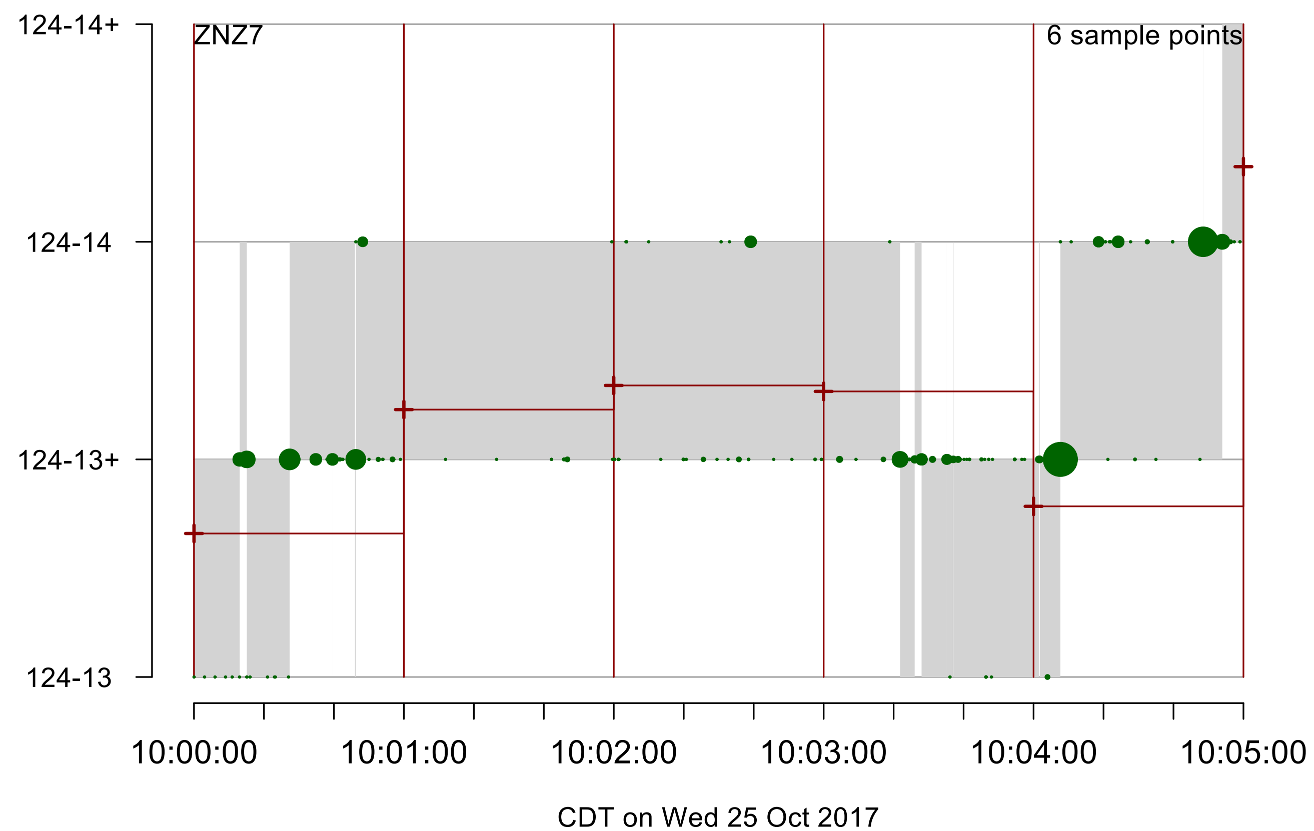


Regular sampling

Quote midpoint



Microprice



Pros and cons of these sampling methods

- Trade times measure market activity
- Trades are often split, repeated trades

  reason to exclude successive ones at same price
- Quote midpoints have less reversion than trades
- Microprice not really effective at reducing noise
- Final selection depends on how you use it

Tick size effects

- Nondimensional parameters
- "Large tick" vs "small tick"
- Reversion in general
- Reversion models: Robert/Rosenbaum (2011)

Nondimensional number

- How to compare different products
  - 5000 US stocks
  - US and international stocks
  - 100's of different futures on CME and worldwide
- Natural Gas is "different" than 2-year Treasury

Dimensional parameters

- Have units in them

volatility: price change per √time

$ for stock,
maybe nondim for futures

second, hour, or day

daily volume: shares or lots per day

dimensional

dimensional

Comparison

- Stock A: 1MM shares per day
- Stock B: 2MM shares per day

Is trading 5,000 shares of stock A in one day the "same" as trading 10,000 shares of stock B in one day?

Is trading 10,000 shares of stock B in one day the "same" as trading 10,000 shares of stock A in two days?

Nondimensionalization

Example: market impact
Trade X shares in time T
Price impact I = ?

Nondimensional
exponent and coefficient

$$\frac{I}{\sigma} = \alpha \left(\frac{X}{V} \right)^k$$

Impact as fraction
of daily volatility

Trade size as fraction
of daily volume

Nondimensionalization brings to same scale
Tick size is nondimensional difference

Nondimensional properties

- number of trades before quotes change
- shape of order book across levels
- number of different limit orders
- fraction of time bid-ask spread is 1 tick
- number of price changes per long-term change
- average quote size / average trade size
- Reversion
- *etc*

Tick size spectrum

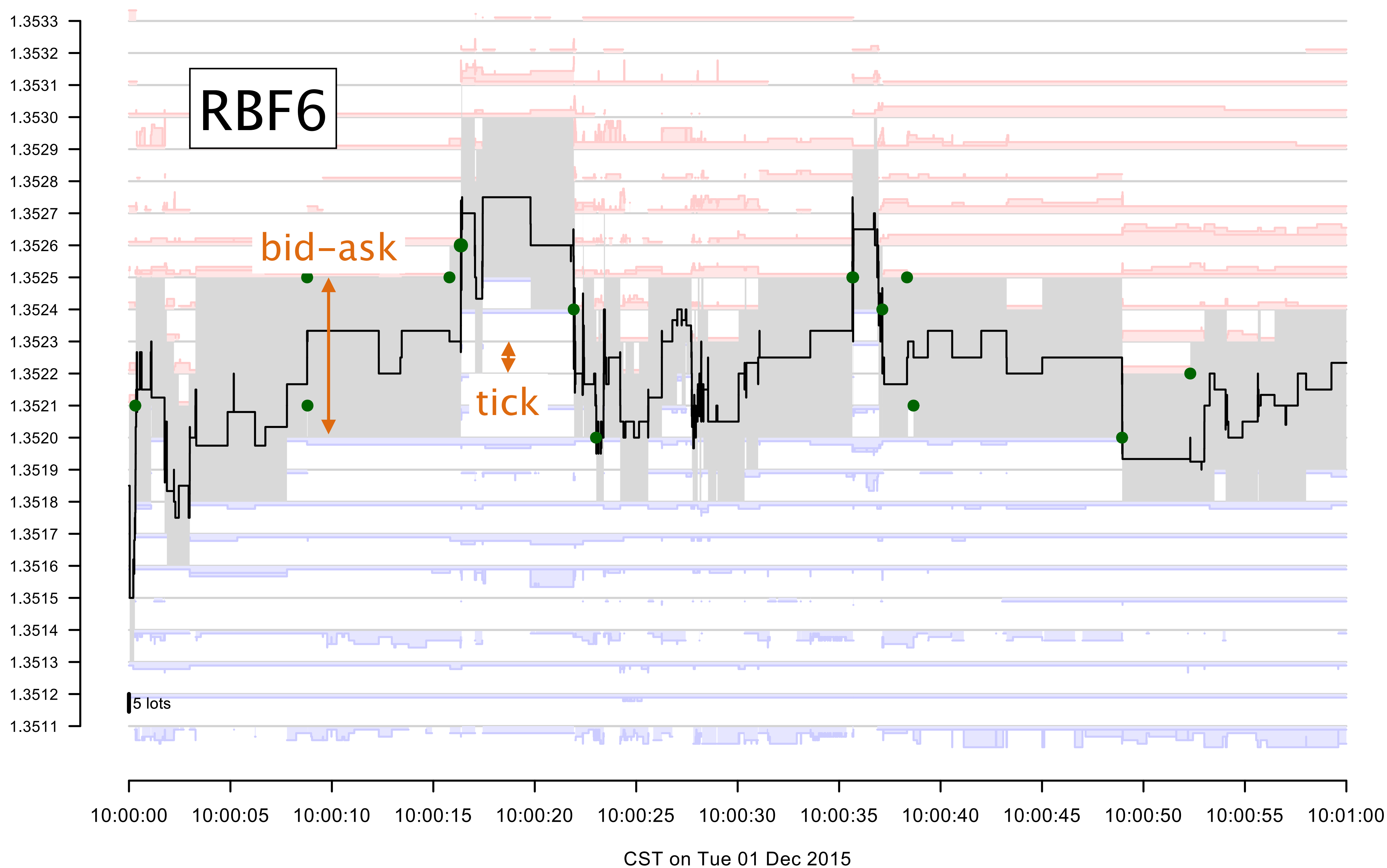
Large tick:
- bid–ask = minimum tick
- quote size $\gg$ trade size
- lots of reversion
- deep order book

$\longleftrightarrow$

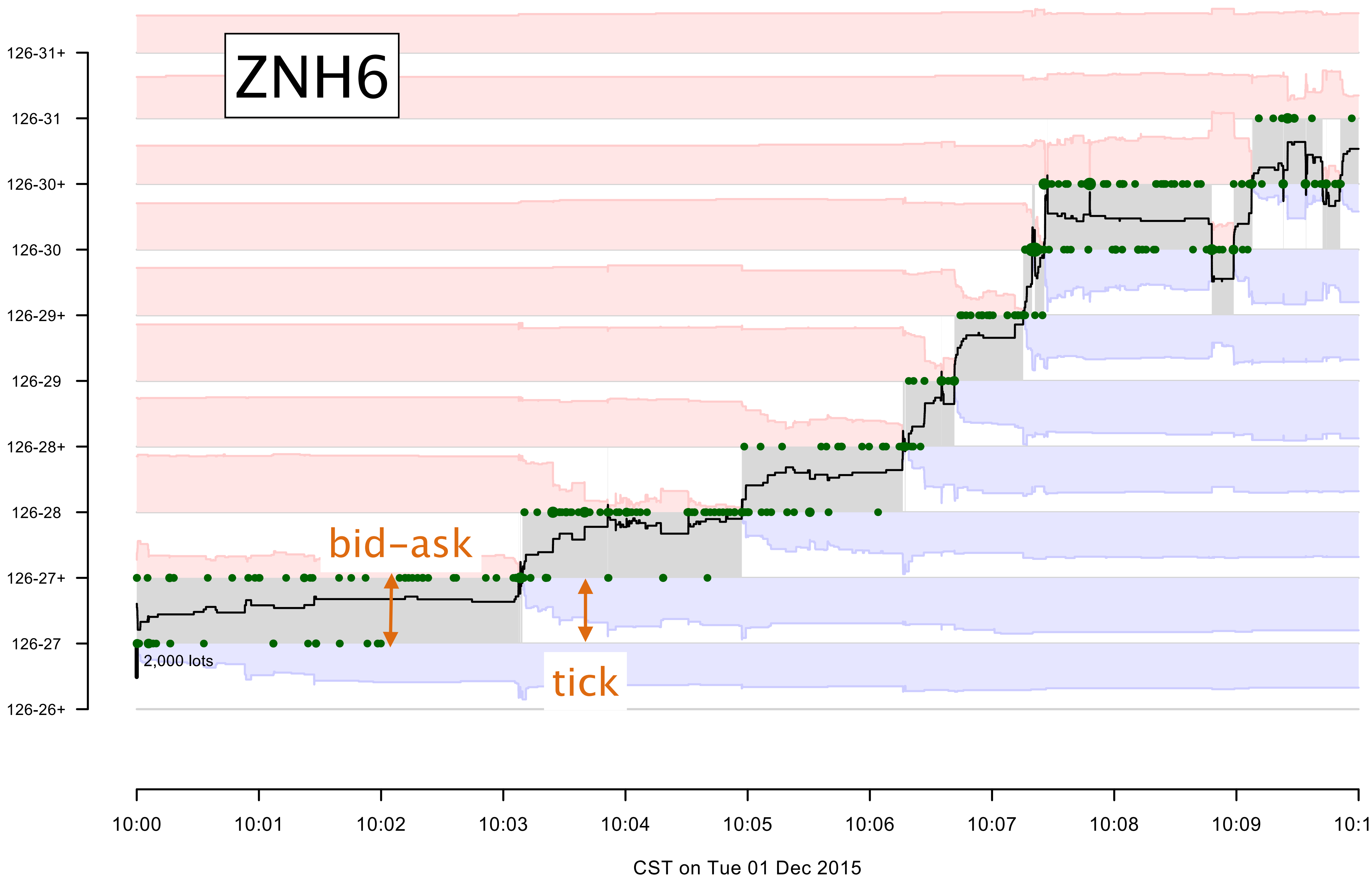
Tick size (relative to intrinsic dynamics)

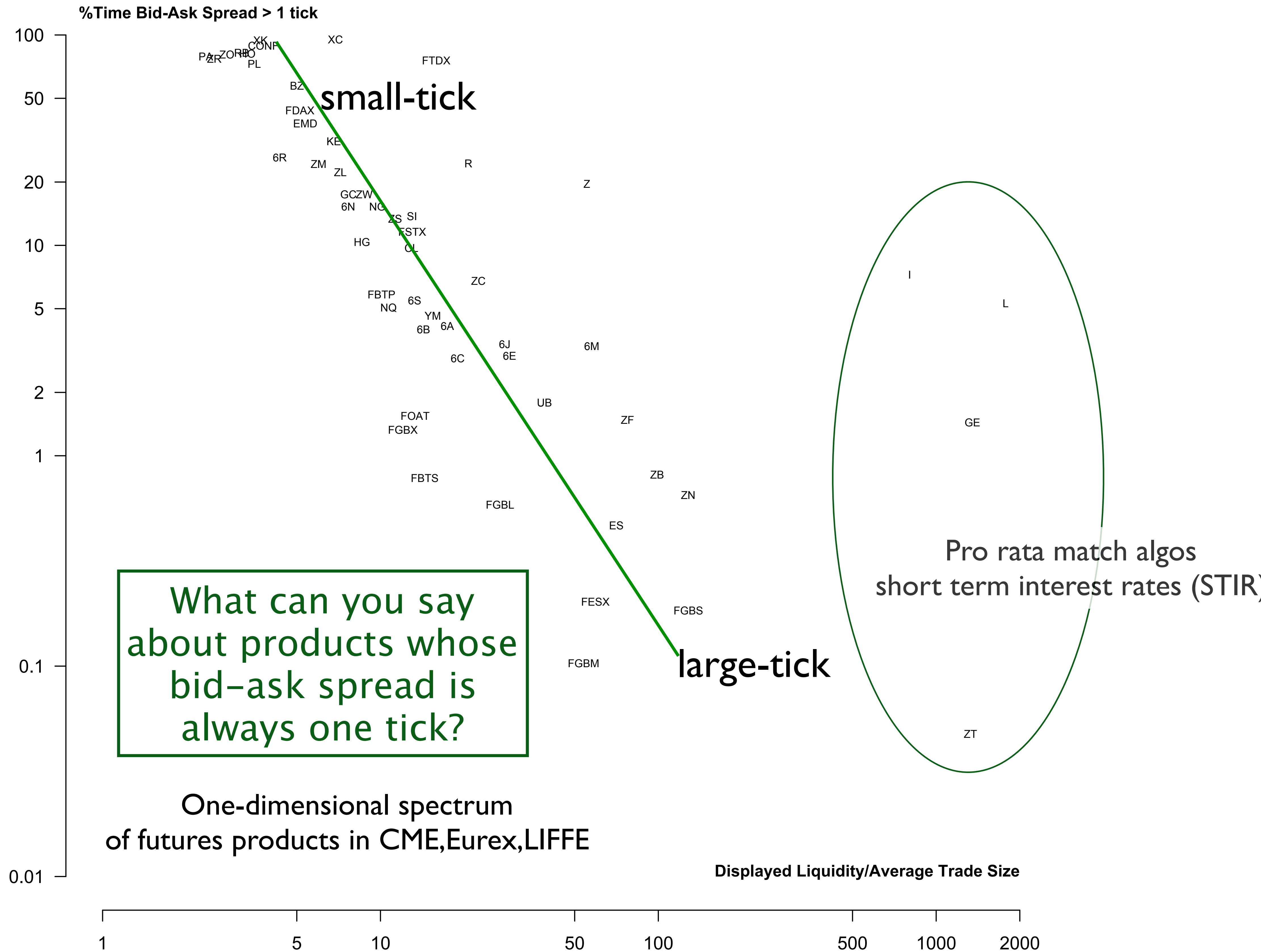
Small tick:
- bid–ask > minimum tick
- quote size ~ trade size
- little reversion
- shallow order book

"small-tick" asset: RBOB Gasoline



"large-tick" asset: 10-year Treasury





What can you say about products whose bid–ask spread is always one tick?

One-dimensional spectrum
of futures products in CME,Eurex,LIFFE

Is tick too large, too small, or just right?

- Spread > minimum indicates that tick may be too small
- Spread = minimum indicates that tick may be too large
- But many products have spread = minimum there are many degrees of "too large"
- Would be nice to have simple measure based only on trade data

Reversion

- Rapid back-and-forth price moves
- Reversion of trade prices

  bid-ask bounce

  large-tick effects
- Reversion of quote midpoint

Roll model (1984)
reversion of trade prices gives effective spread

THE JOURNAL OF FINANCE • VOL. XXXIX, NO. 4 • SEPTEMBER 1984

A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market

RICHARD ROLL*

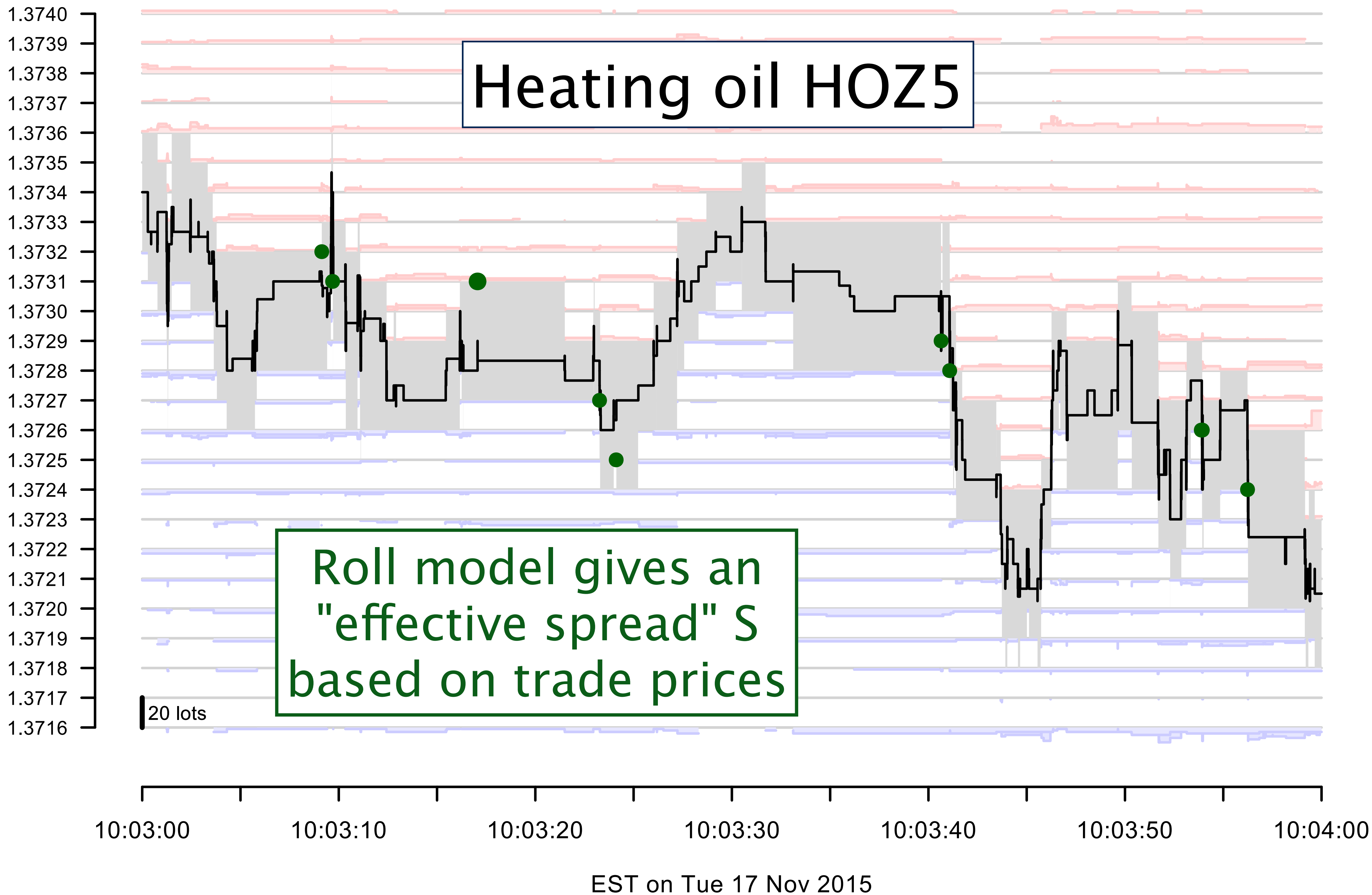
ABSTRACT

In an efficient market, the fundamental value of a security fluctuates randomly. However, trading costs induce negative serial dependence in successive observed market price changes. In fact, given market efficiency, the effective bid-ask spread can be measured by

$$\text{Spread} = 2\sqrt{-\text{cov}}$$

where "cov" is the first-order serial covariance of price changes. This implicit measure of the bid-ask spread is derived formally and is shown empirically to be closely related to firm size.

Average spread when not 1 tick



Journal of Financial Econometrics, 2011, Vol. 9, No. 2, 344–366

Uncertainty zone model

A New Approach for the Dynamics of Ultra-High-Frequency Data: The Model with Uncertainty Zones

Christian Y. Robert
CREST and ENSAE Paris Tech

Mathieu Rosenbaum
CMAP-École Polytechnique Paris UMR CNRS 7641

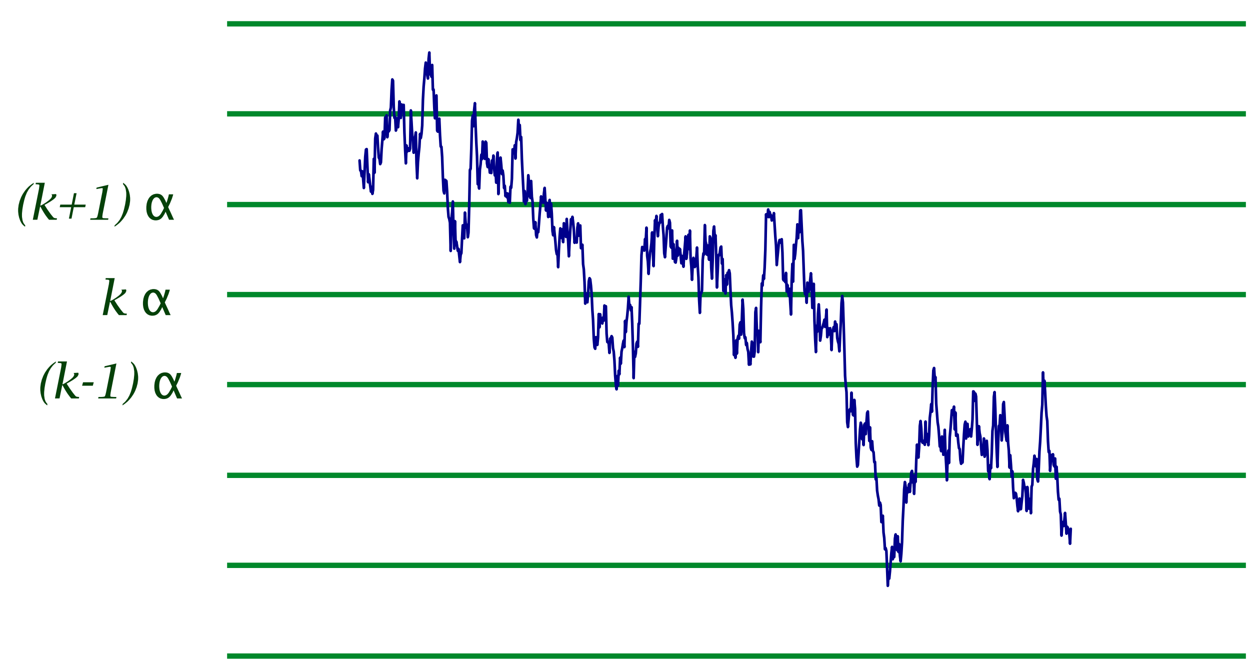
ABSTRACT
In this paper, we provide a model which accommodates the assumption of a continuous efficient price with the inherent properties of ultra-high-frequency transaction data (price discreteness, irregular temporal spacing, diurnal patterns...). Our approach consists in designing a stochastic mechanism for deriving the transaction prices from the latent efficient price. The main idea behind the model is that, if a transaction occurs at some value on the tick grid and leads to a price change, then the efficient price has been close enough to this value shortly before the transaction. We call uncertainty zones the bands around the mid-tick grid where the efficient price is too far from the tick grid to trigger a price change. In our setting, the width of these uncertainty zones quantifies the aversion to price changes of the market participants. Furthermore, this model enables us to derive approximated values of the efficient price at some random times, which is particularly useful for building statistical procedures. Convincing results are obtained through a simulation study and the use of the model over 10 representative stocks.

Theoretical model for tick size and reversion

"True" underlying price (or log-price) *X(t)*:

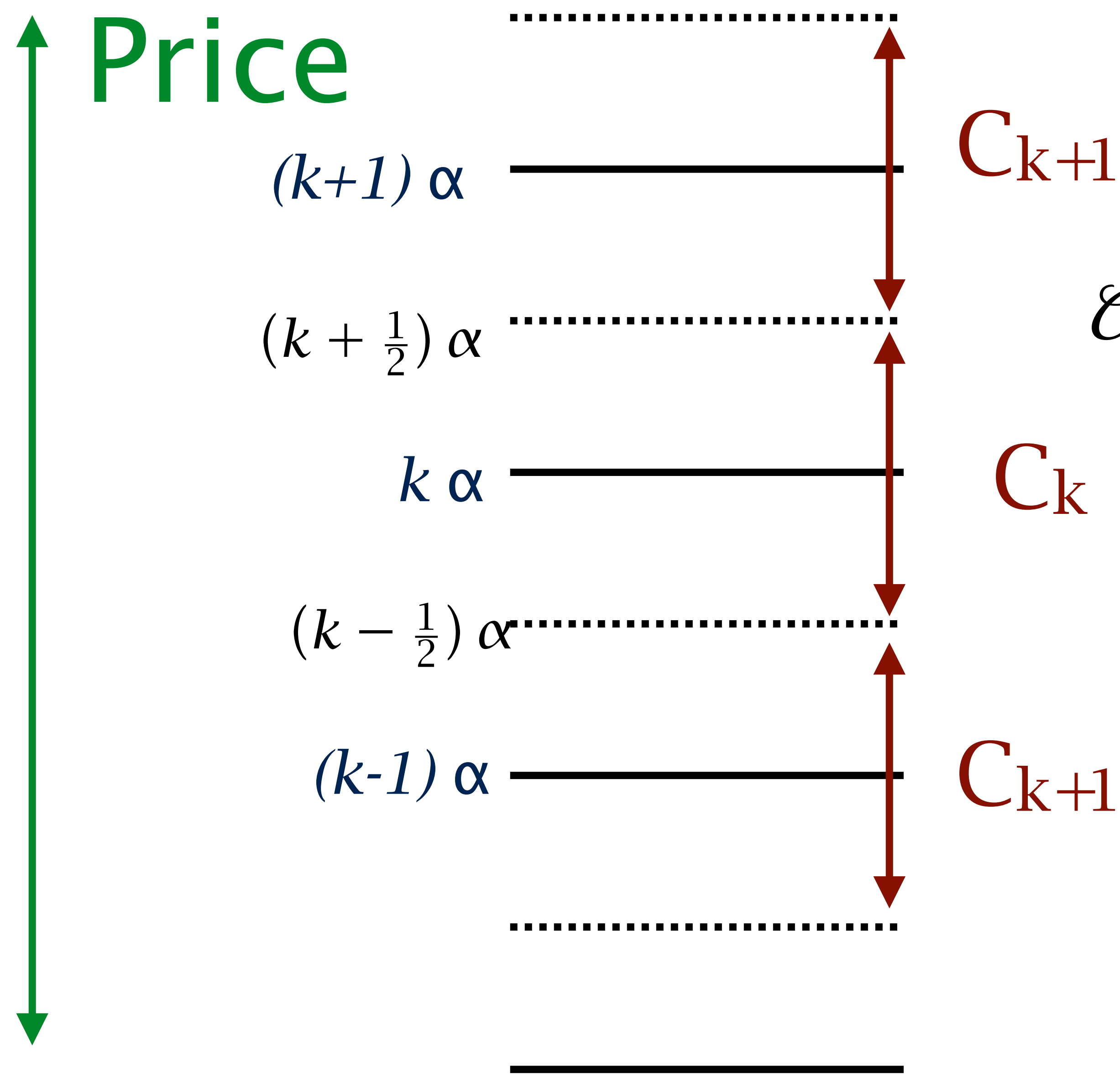
$$dX = \sigma(t)\, dW(t)$$

How does the continuous process *X(t)* relate to the discrete grid of spacing **α**?



What does not work

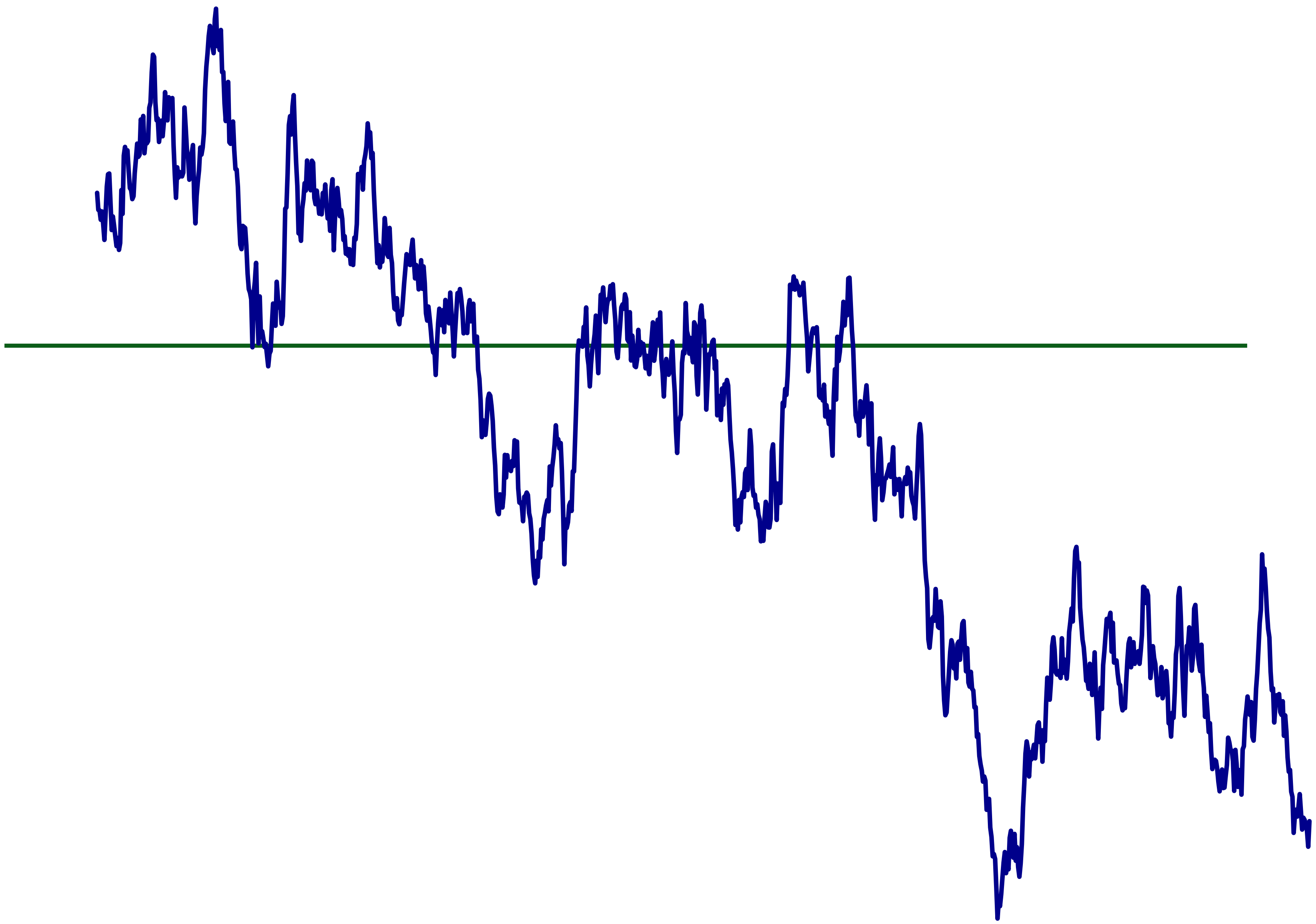
Trades happen at kα while price in C_k



$$\mathcal{C}_k \;=\; \left\{ (k-\tfrac{1}{2})\alpha \;<\; X_t \;<\; (k+\tfrac{1}{2})\alpha \right\}$$

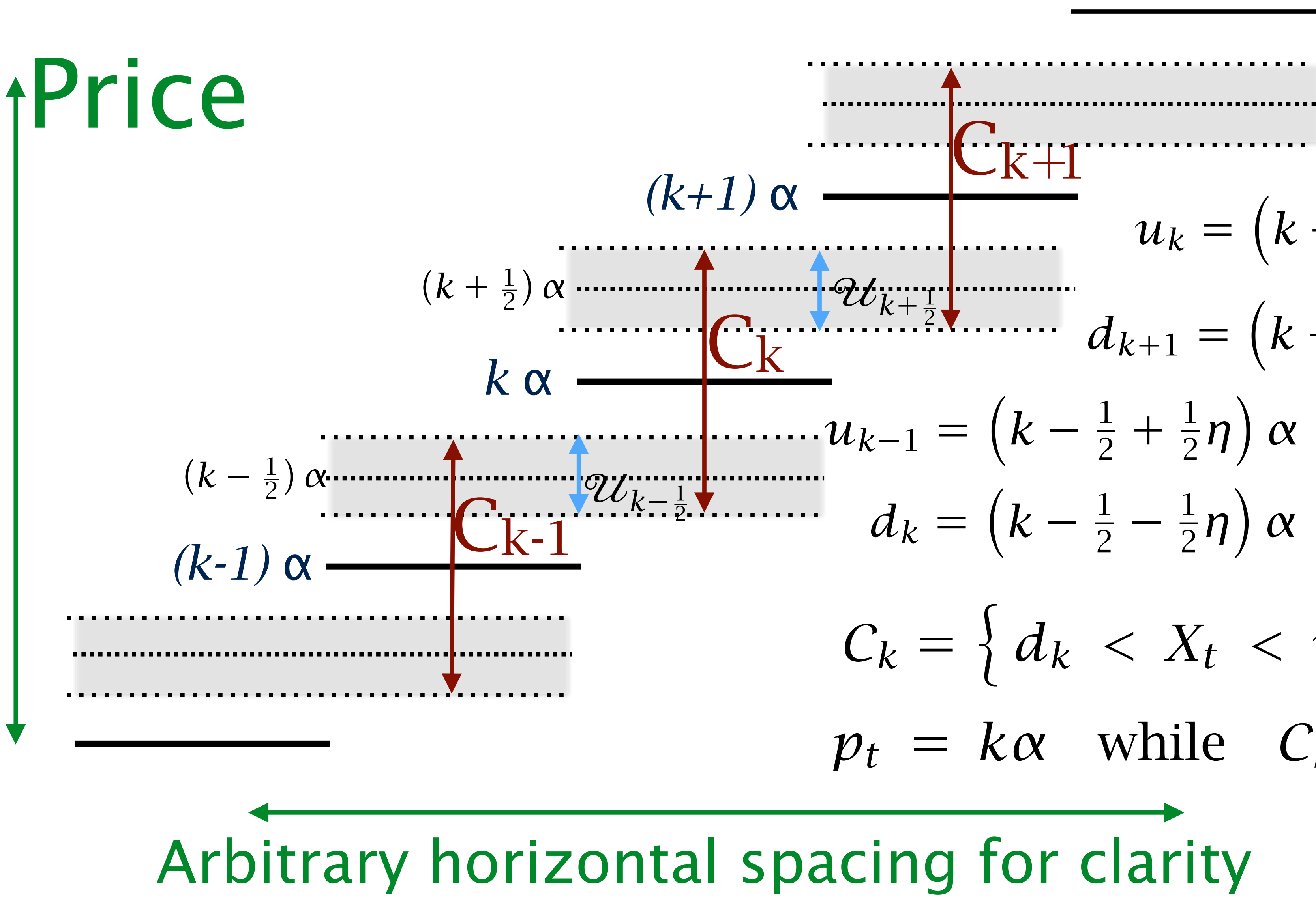
$$P_t = k\alpha \quad \text{while} \quad X_t \in \mathcal{C}_k$$

Does not work because infinite fluctuations across boundaries



Uncertainty zones

$0 < \eta < 2$ This picture $0<\eta<1$



$$u_k = \left(k + \tfrac{1}{2} + \tfrac{1}{2}\eta\right)\alpha$$

$$d_{k+1} = \left(k + \tfrac{1}{2} - \tfrac{1}{2}\eta\right)\alpha$$

$$u_{k-1} = \left(k - \tfrac{1}{2} + \tfrac{1}{2}\eta\right)\alpha$$

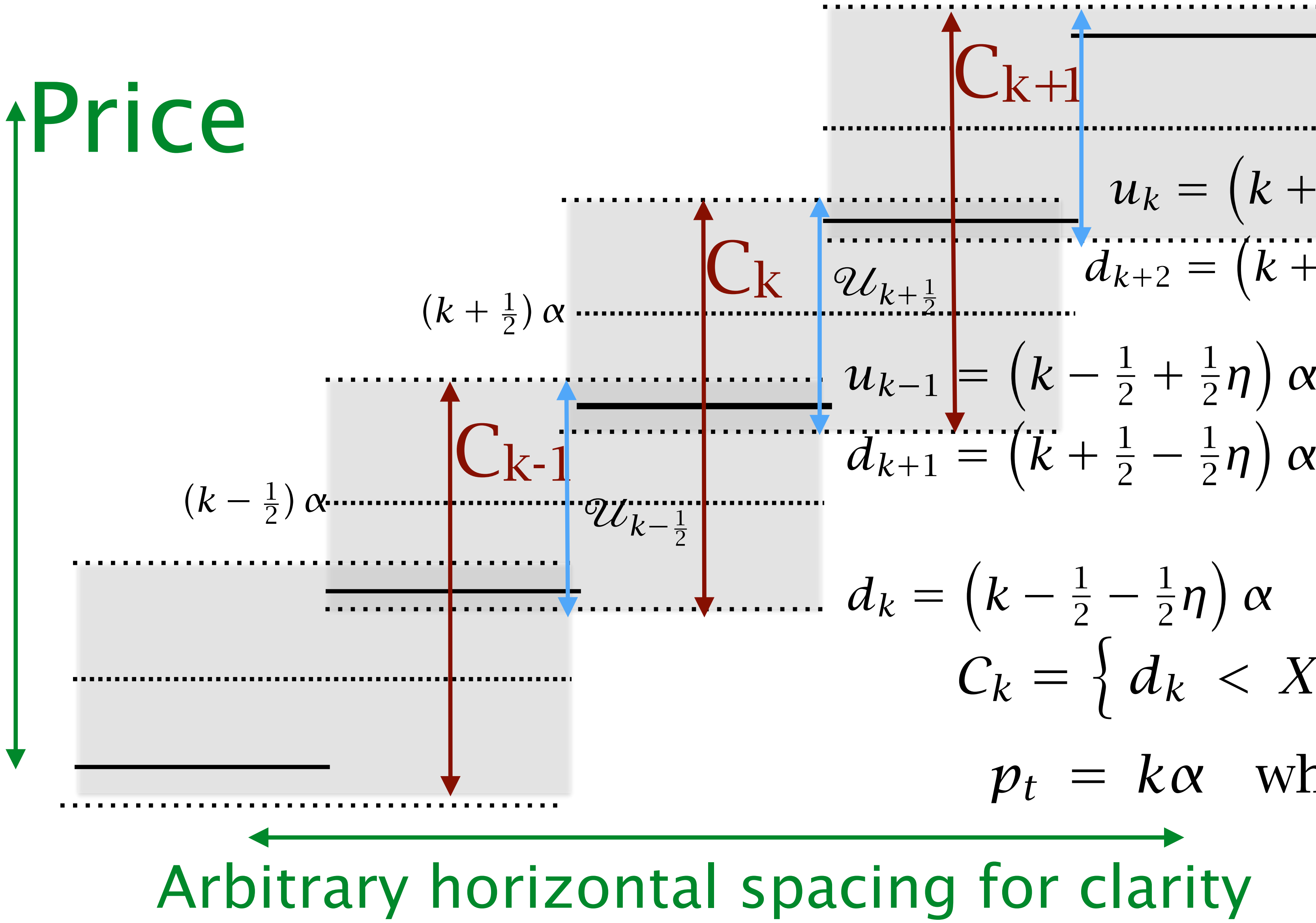
$$d_k = \left(k - \tfrac{1}{2} - \tfrac{1}{2}\eta\right)\alpha$$

$$C_k = \left\{ d_k < X_t < u_k \right\}$$

$$p_t = k\alpha \quad \text{while} \quad C_k$$

Uncertainty zones

$0 < \eta < 2$ This picture $1<\eta<2$



$$u_k = \left(k + \tfrac{1}{2} + \tfrac{1}{2}\eta\right)\alpha$$

$$d_{k+2} = \left(k + 1 + \tfrac{1}{2} - \tfrac{1}{2}\eta\right)\alpha$$

$$u_{k-1} = \left(k - \tfrac{1}{2} + \tfrac{1}{2}\eta\right)\alpha$$

$$d_{k+1} = \left(k + \tfrac{1}{2} - \tfrac{1}{2}\eta\right)\alpha$$

$$d_k = \left(k - \tfrac{1}{2} - \tfrac{1}{2}\eta\right)\alpha$$

$$C_k = \left\{ d_k < X_t < u_k \right\}$$

$$p_t = k\alpha \quad \text{while} \quad C_k$$

- Trades P_t occur at price $k\,\alpha$ while $X_t \in C_k$
- At least one trade occurs during each interval
- Trade price changes happen when P_t exits C_k
- Parameter η measures "stickiness" of price how far "real" price has to move beyond mid before trade prices adjust
- $\eta<1$: trade at new price before X_t gets there

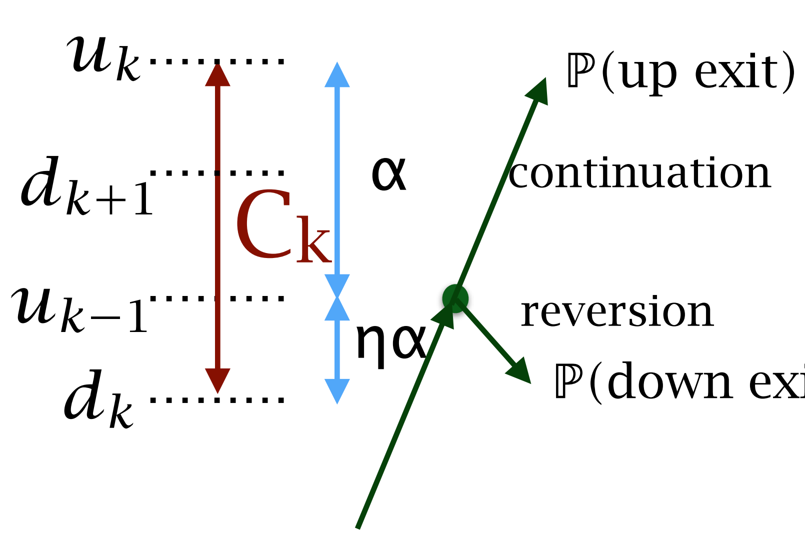
  tick size is "too large"
- $\eta>1$: trade at new price after X_t gets there

  tick size is "too small"
- "Ideal" market would have $\eta = 1$

How to estimate **η**?

Observe trades at P_k, previous trades were at $P_{k\text{-}1}$

Which direction will it exit next?



$$\mathbb{P}(\text{up exit}) = \frac{\eta}{1+\eta}$$

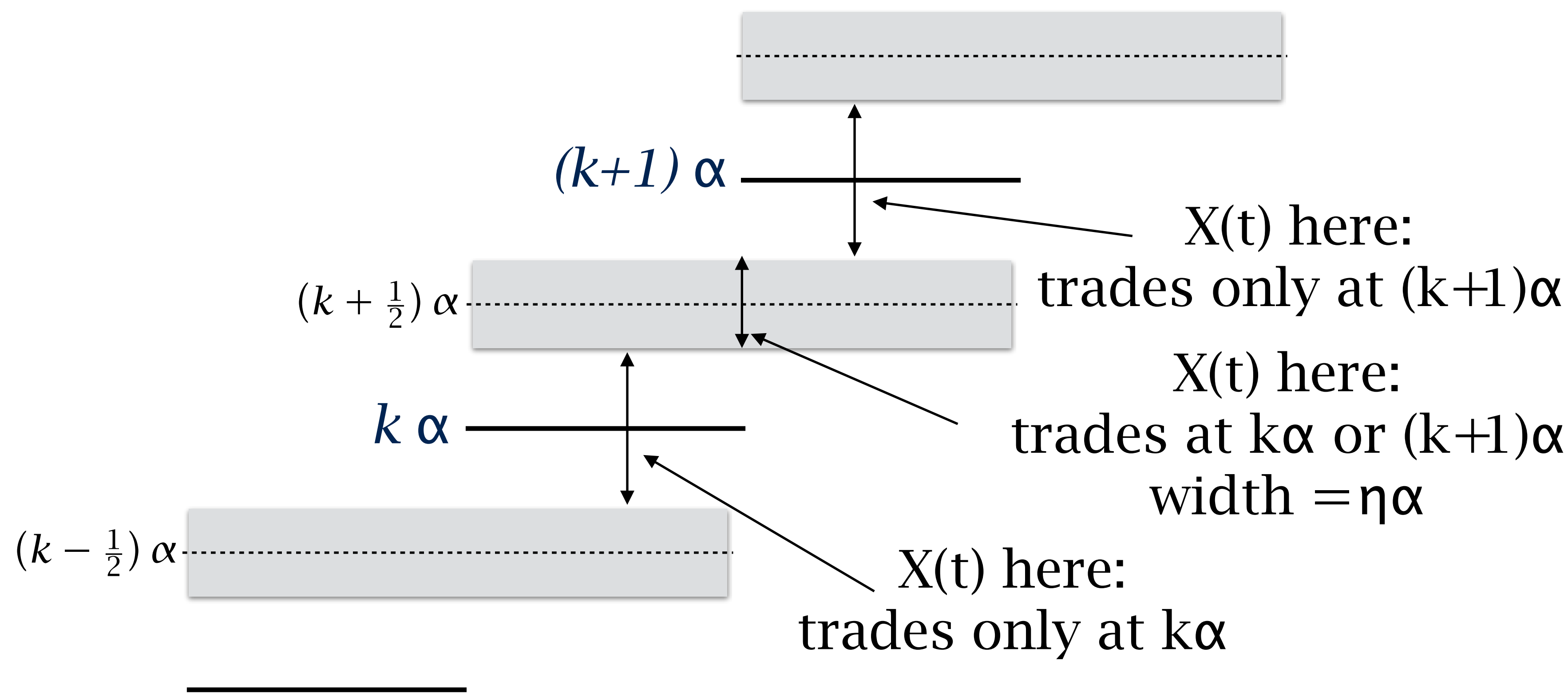
$$\frac{\#\text{ continuations}}{\#\text{ reversions}} = \eta$$

$$\mathbb{P}(\text{down exit}) = \frac{1}{1+\eta}$$

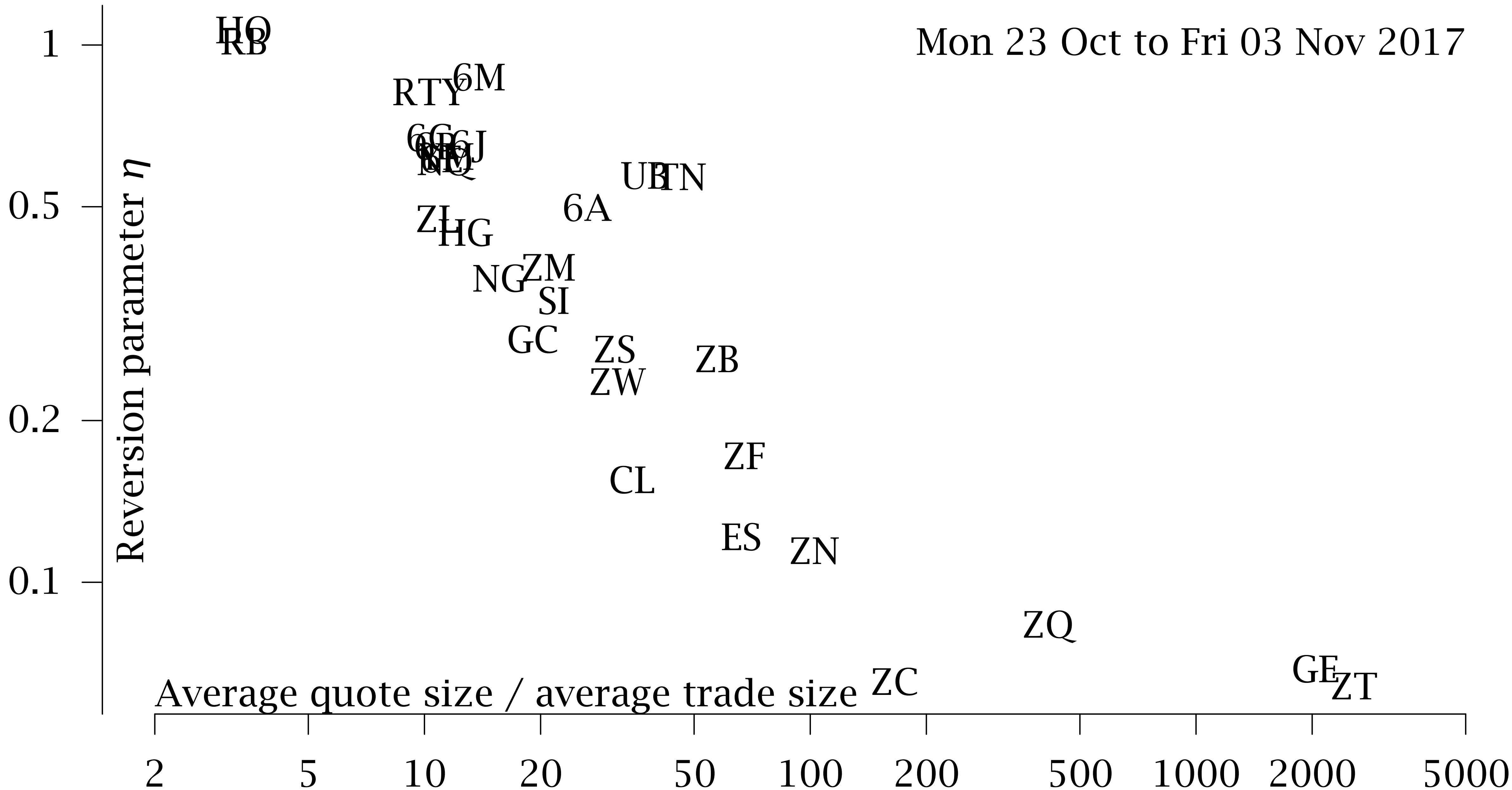
Serial correlation of price changes $\quad \rho = -\dfrac{1-\eta}{1+\eta}$

Ignore size of price changes
Look only at direction
Works best when all moves ±1

Effective spread = $\eta\ \alpha$



Correlation with other variables



Can you replace trade prices by quote midpoints?

- In principle, yes

  should eliminate bid-ask bounce

- In practice, difficulties with spread widening

  can restrict to times spread = 1 tick

  but this does not work for small-tick assets

Predict effects of tick size change

Market Microstructure and Liquidity
Vol. 1, No. 1 (2015) 1550003 (29 pages)
© World Scientific Publishing Company
DOI: 10.1142/S2382626615500033

Large Tick Assets: Implicit Spread and Optimal Tick Size

Khalil Dayri
Algorithms and Analytics
Bloomberg Tradebook, United Kingdom
kdayri@bloomberg.net

Mathieu Rosenbaum*
Laboratory of Probability and Random Models
University Pierre and Marie Curie (Paris 6), France
mathieu.rosenbaum@upmc.fr

In this work, we provide a framework linking microstructural properties of an asset to the tick value of the exchange. In particular, we bring to light a quantity, referred to as *implicit spread*, playing the role of spread for large tick assets, for which the effective spread is almost always equal to one tick. The relevance of this new parameter is shown both empirically and theoretically. This implicit spread allows us to quantify the tick sizes of large tick assets and to define a notion of *optimal tick size*. Moreover, our results open the possibility of forecasting the behavior of relevant market quantities after a change in the tick value and to give a way to modify it in order to reach an optimal tick size. Thus, we provide a crucial tool for regulators and trading platforms in the context of high frequency trading.

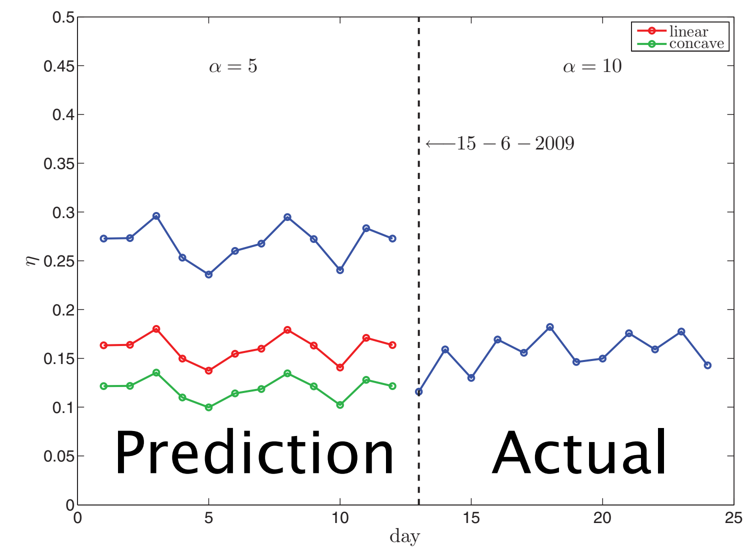


Fig. 5. Testing the prediction of η on the Bobl futures. The blue lines show the daily measures of η. The red and green lines are the daily predictions associated to the future tick value.

Therefore, under reasonable assumptions, we are able to forecast the value of η after a change in the tick value. In order to check these formulas on real data, we use the Bobl contract. The tick value of this asset has been multiplied by two on 2009, June 15. For 12 trading days before 2009, June 15, we give in Fig. 5 the estimates of the value of η after the change of the tick value given by Eq. (5) (version 1 above) with $\beta = 1$ and $\beta = 1/2$.

How to predict the consequences of a tick value change?
Evidence from the Tokyo Stock Exchange pilot program

Weibing Huang[1], Charles-Albert Lehalle[2] and Mathieu Rosenbaum[1]

July 27, 2015

Abstract

The tick value is a crucial component of market design and is often considered the most suitable tool to mitigate the effects of high frequency trading. The goal of this paper is to demonstrate that the approach introduced in Dayri and Rosenbaum (2015) allows for an ex ante assessment of the consequences of a tick value change on the microstructure of an asset. To that purpose, we analyze the pilot program on tick value modifications started in 2014 by the Tokyo Stock Exchange in light of this methodology. We focus on forecasting the future cost of market and limit orders after a tick value change and show that our predictions are very accurate. Furthermore, for each asset involved in the pilot program, we are able to define (ex ante) an optimal tick value. This enables us to classify the stocks according to the relevance of their tick value, before and after its modification.

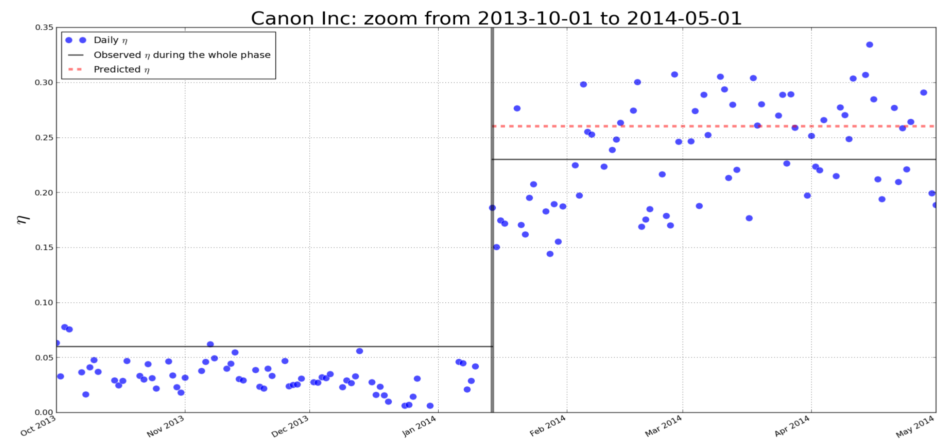


Figure 3: Daily estimations of η for the last 3.5 months of Phase 0 and the first 3.5 months of Phase 1 for the stock Canon Inc.

To end this subsection, as an illustration, we give in Figure 3 detailed results about η for the stock Canon Inc. More precisely, we provide daily estimations of η for the last 3.5 months of Phase 0 and the first 3.5 months of Phase 1. We also add the average values of η during both phases together with our forecast for the value of η in Phase 1. We see on this example that our prediction is very close to the realized value.

Application to algorithms

1. Reversion

- Reversion = likelihood of successive moves in opposite directions
- Trade price reversion:

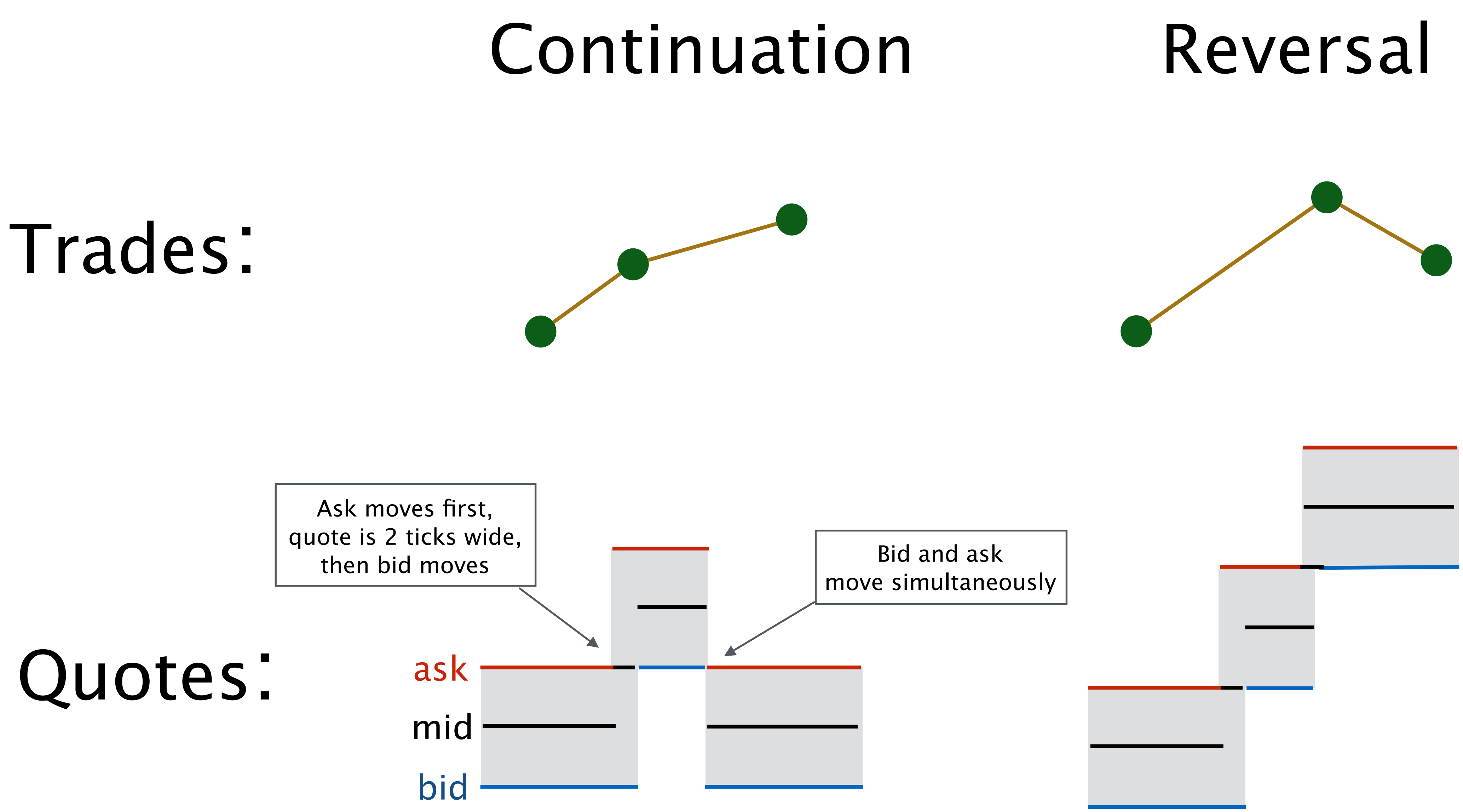
Bid-ask bounce (Roll model)

Uncertainty zones (Robert/Rosenbaum/Dayri)

- Quote price reversion:

Uncertainty zones/large-tick effects

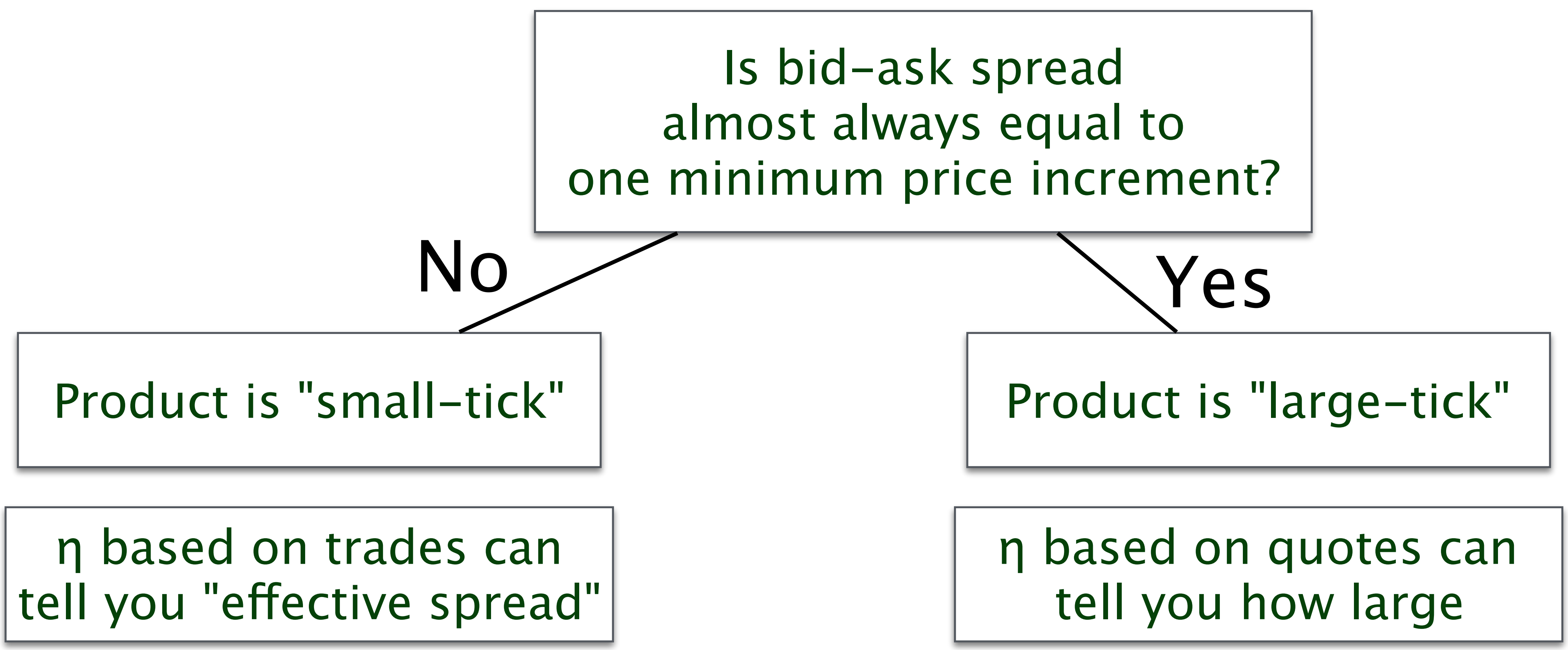
Multi-tick markets



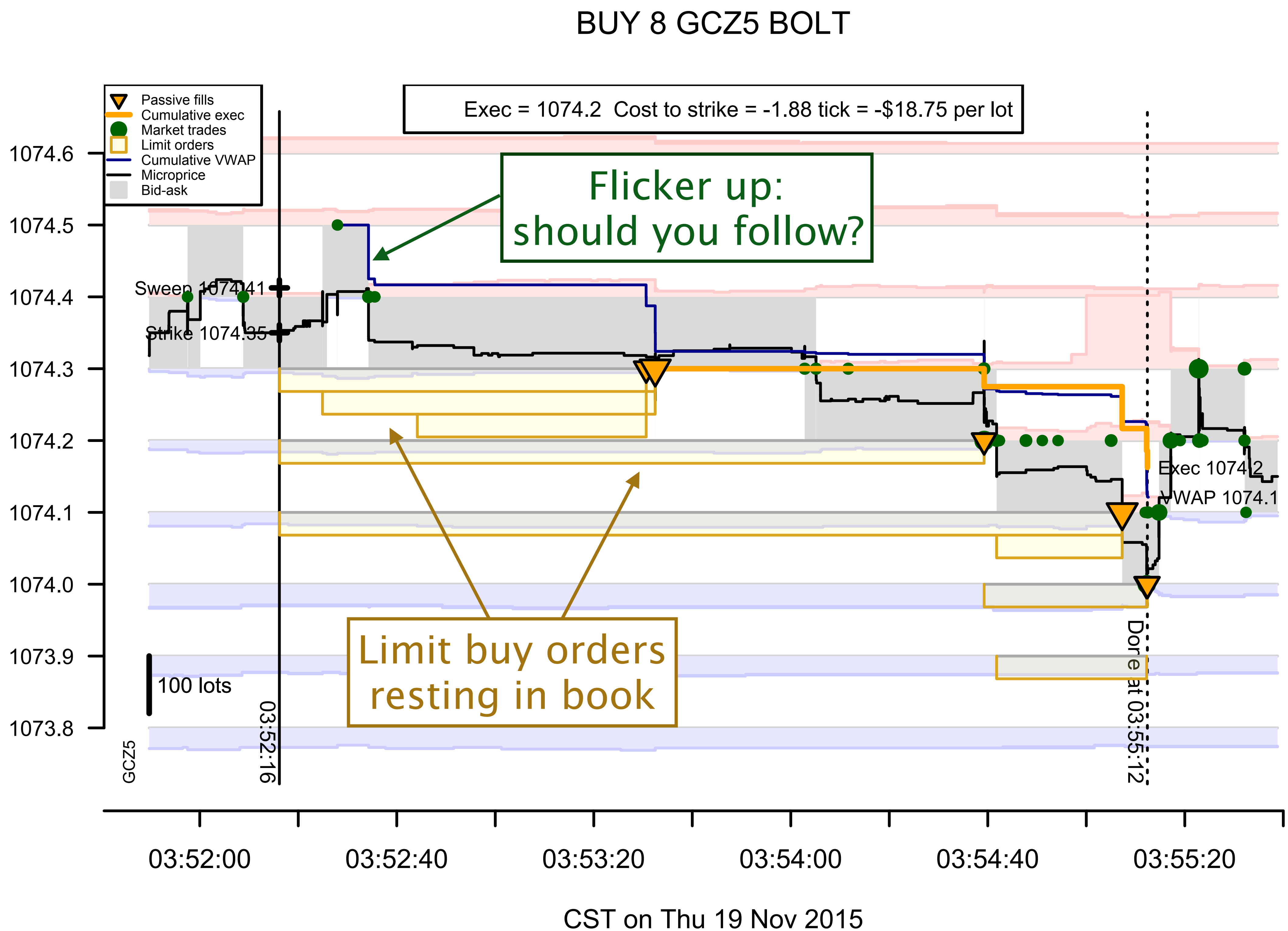
$$\eta = \frac{\#\ \text{Continuations}}{\#\ \text{Reversals}}$$

Significance of η: large tick

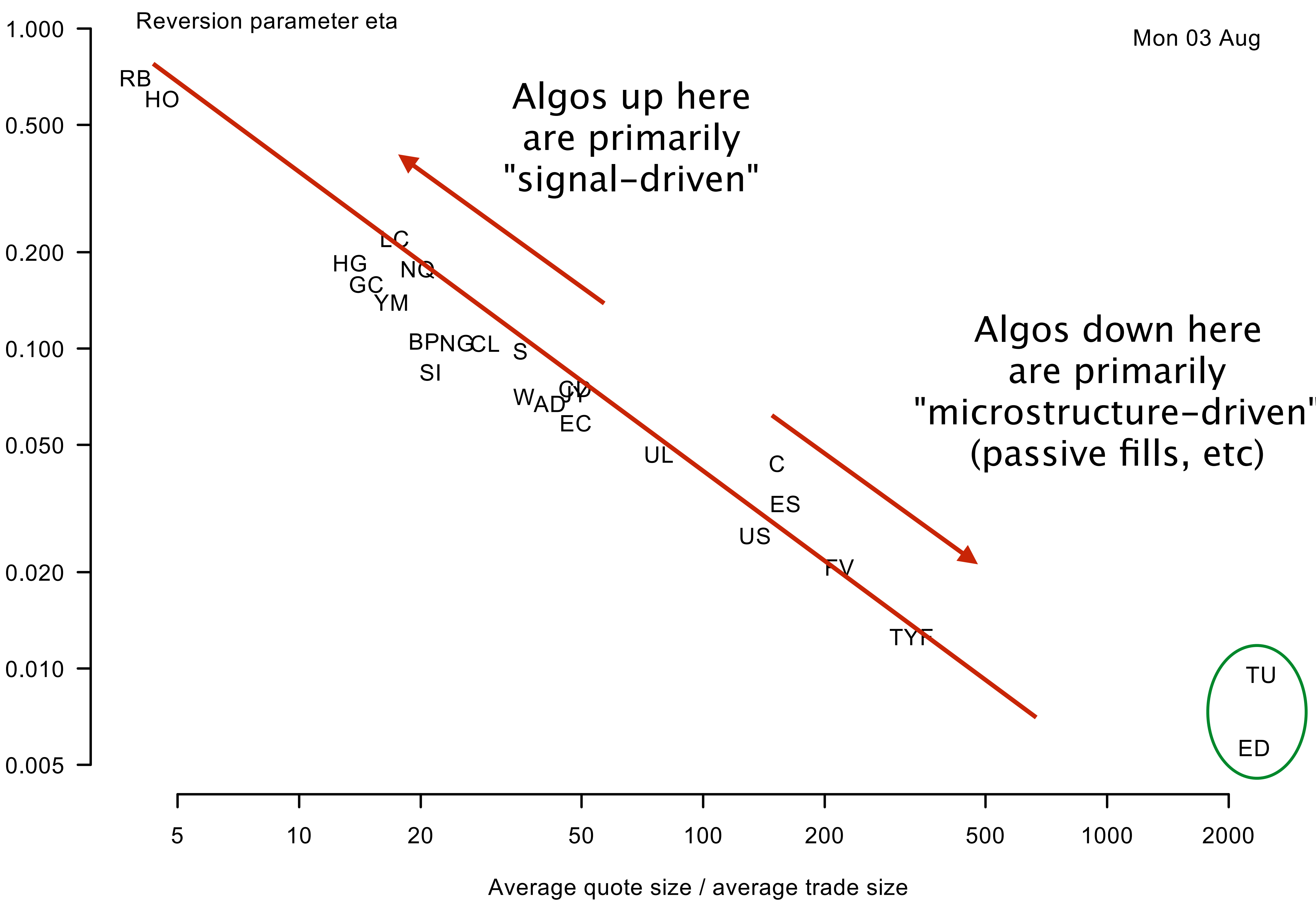
For a particular product:



Reversion complicates trading



Spectrum based on reversion parameter (using filtered quote midpoint changes)



Reversion complicates volatility estimation

- Definition: volatility is (the square root of) variance of price changes per unit time on times long enough to ignore microstructure.
- Want to use data as high-resolution as possible to get better statistical estimation properties
- If short-term price moves are serially correlated, then volatility estimated from short-term moves will be larger or smaller than long-time volatility.
- Simple solution: 1-minute sampling as for equities
- Can we do better?