PRINCETON UNIVERSITY

ORF 474: High Frequency Trading
Spring 2020
Robert Almgren

Lecture 2b

Feb 12, 2020

Today

- Machine precision and timestamps
- Matching trades and quotes
- Bid-ask spread
- Volatility

Machine precision

Q: Can R represent nanoseconds?

A: Yes, if you are careful.

and financial data scientist!

What Every Computer Scientist Should Know About Floating-Point Arithmetic

DAVID GOLDBERG

Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304

ACM Computing Surveys, Vol 23, No 1, March 1991

Floating-point arithmetic is considered an esotoric subject by many people. This is rather surprising, because floating-point is ubiquitous in computer systems: Almost every language has a floating-point datatype; computers from PCs to supercomputers have floating-point accelerators; most compilers will be called upon to compile floating-point algorithms from time to time; and virtually every operating system must respond to floating-point exceptions such as overflow This paper presents a tutorial on the aspects of floating-point that have a direct impact on designers of computer systems. It begins with background on floating-point representation and rounding error, continues with a discussion of the IEEE floating-point standard, and concludes with examples of how computer system builders can better support floating point.

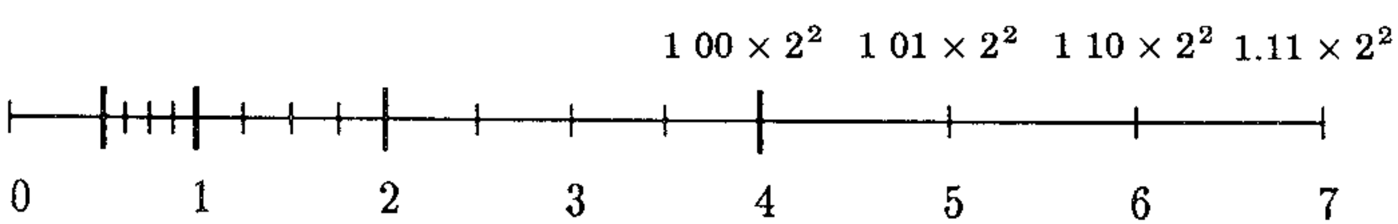
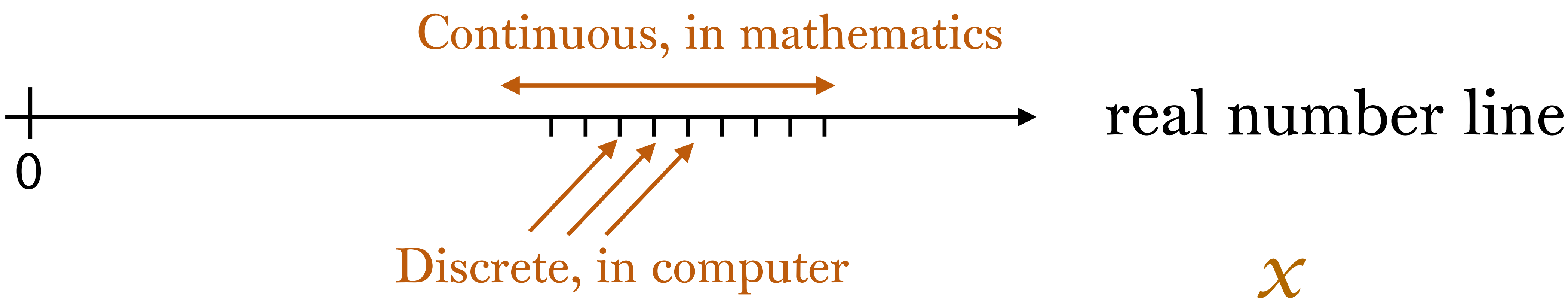
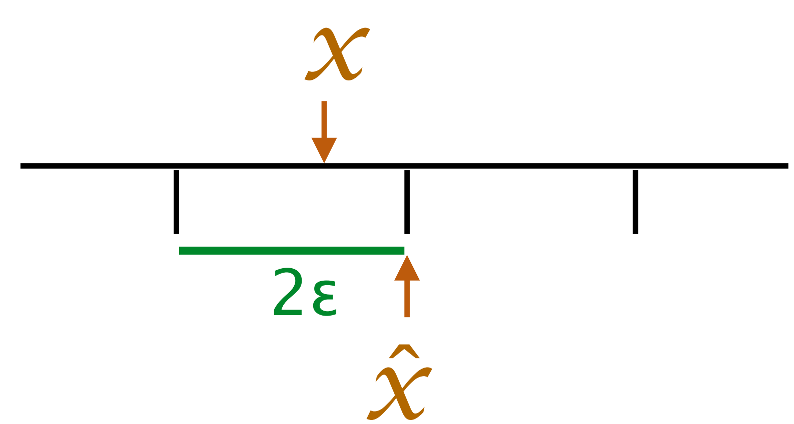


Figure 1. Normalized numbers when $\beta = 2$, $p = 3$, $e_{\min} = -1$, $e_{\max} = 2$.

Machine precision





$$\hat{x} = x \cdot (1 + \delta)$$

Machine number True number

$$\hat{x} - x = x \cdot \delta$$

$$|\delta| \leq \epsilon$$

Machine precision: $\boldsymbol{\varepsilon} = 10^{-16}$
for IEEE double precision floating point

Date and time separately:

Date = integer number of days since 1970-01-01
Time = double precision seconds since midnight
(New York local time)

$$
\begin{aligned}
|\hat{x} - x| &\le x \cdot \epsilon \\
&\le 24 \cdot 60 \cdot 60 \,\text{sec} \cdot 10^{-16} \\
&\approx 10^{-11}\,\text{sec} \\
&= 0.01\,\text{nsec}
\end{aligned}
$$

Yes: R times can represent nanoseconds, if stored separately from date (as for TAQ)

TAQ trade file

(Not every TAQ file has full 9 digits)

```
DATE,TIME_M,EX,SYM_ROOT,SYM_SUFFIX,TR_SCOND,SIZE,PRICE,TR_CORR,TR_SEQNU
M,TR_SOURCE,TR_RF
20190102,9:30:00.955640992,Q,FOXF,,@O X,1662,57.81,00,8150,N,
20190102,9:30:00.955654700,Q,FOXF,,@  Q,1662,57.81,00,8151,N,
20190102,9:30:02.783472740,D,FOXF,,@  I,50,57,00,9710,N,N
20190102,9:30:05.163990069,D,FOXF,,@,210,59.465,00,10691,N,Q
20190102,9:30:09.066601950,D,FOXF,,@,150,57.8407,00,11691,N,N
20190102,9:30:32.738968225,P,FOXF,,@F I,2,57.05,00,14115,N,
20190102,9:30:32.738977133,P,FOXF,,@  Q,2,57.05,00,14116,N,
20190102,9:30:32.837962259,V,FOXF,,@  I,1,59.34,00,14120,N,
20190102,9:30:35.821049617,D,FOXF,,@  I,50,57.02,00,14456,N,Q
```

Match to 9 digits

R data.table

```
         date            time exch  sym          prc  siz cond corr   seq src rf
1  2019-01-02 34200.955640992    Q FOXF 57.810000000 1662 @O X    0  8150   N
2  2019-01-02 34200.955654700    Q FOXF 57.810000000 1662 @  Q    0  8151   N
3  2019-01-02 34202.783472740    D FOXF 57.000000000   50 @  I    0  9710   N  N
4  2019-01-02 34205.163990069    D FOXF 59.465000000  210    @    0 10691   N  Q
5  2019-01-02 34209.066601950    D FOXF 57.840700000  150    @    0 11691   N  N
6  2019-01-02 34232.738968225    P FOXF 57.050000000    2 @F I    0 14115   N
7  2019-01-02 34232.738977133    P FOXF 57.050000000    2 @  Q    0 14116   N
8  2019-01-02 34232.837962259    V FOXF 59.340000000    1 @  I    0 14120   N
9  2019-01-02 34235.821049617    D FOXF 57.020000000   50 @  I    0 14456   N  Q
10 2019-01-02 34258.966139798    Q FOXF 58.520000000    1 @  I    0 16470   N
```

Date and time in one variable
R POSIXct = double precision seconds
since 1970-01-01 00:00:00 UTC
(need this if trading worldwide)

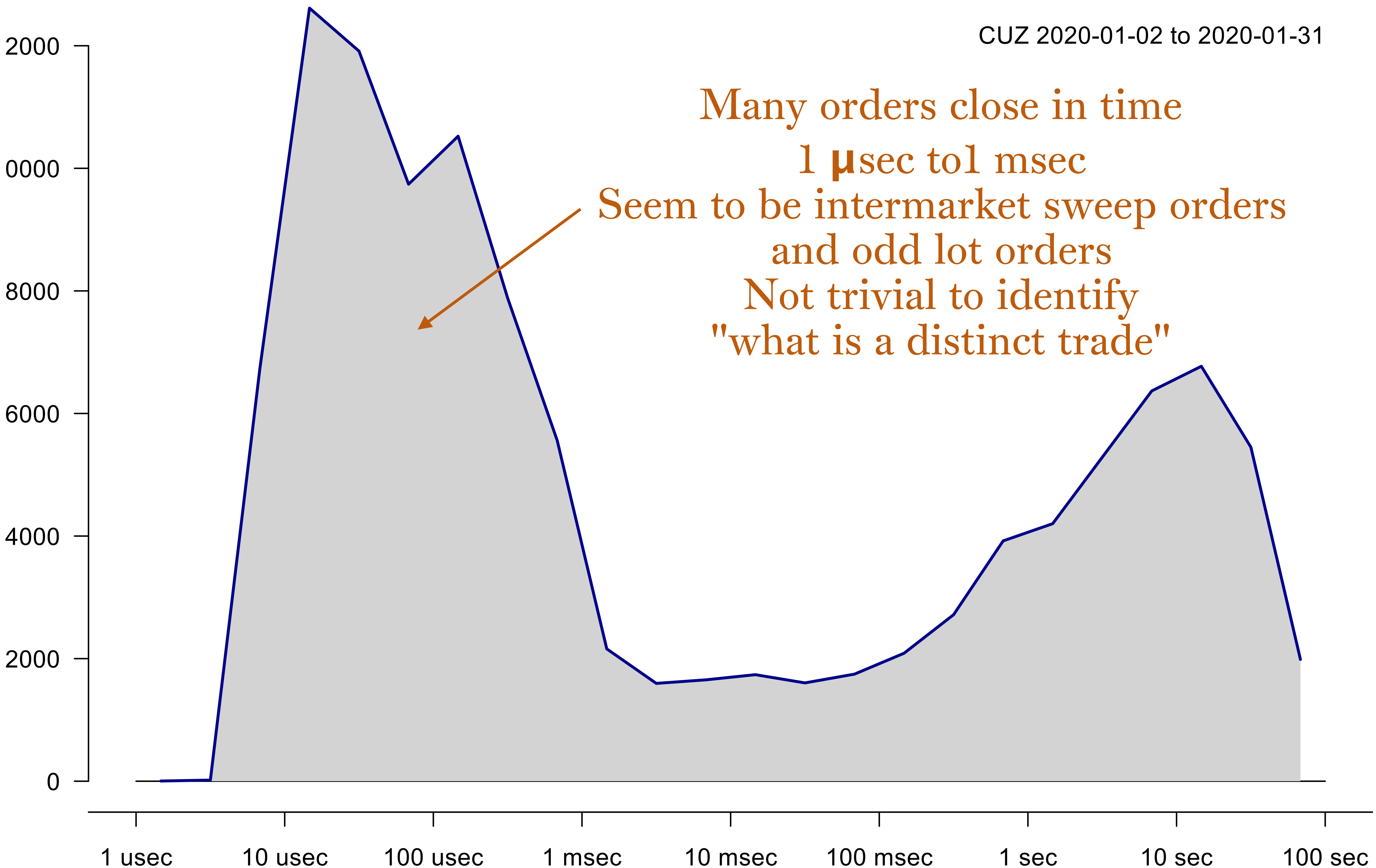
$$
\begin{aligned}
|\hat{x} - x| &\le x \cdot \epsilon \\
&\le (2019 - 1970) \cdot 365 \\
&\quad \cdot 24 \cdot 60 \cdot 60 \cdot 10^{-16} \text{ sec} \\
&\approx 1.5 \times 10^{-7} \text{ sec} \\
&= 150 \text{ nsec}
\end{aligned}
$$

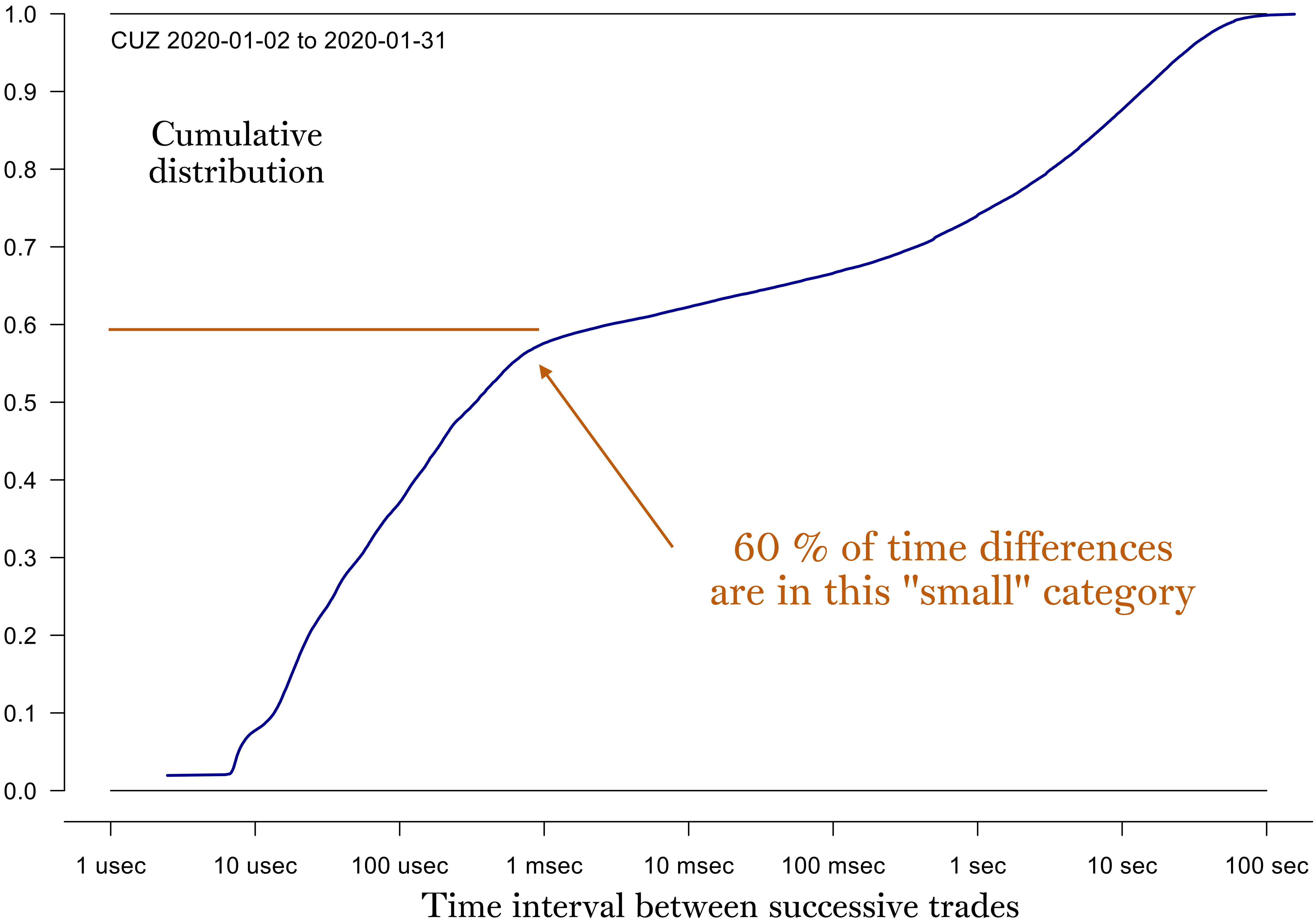
R POSIXct *cannot* represent nanoseconds
Do not convert times via POSIXct

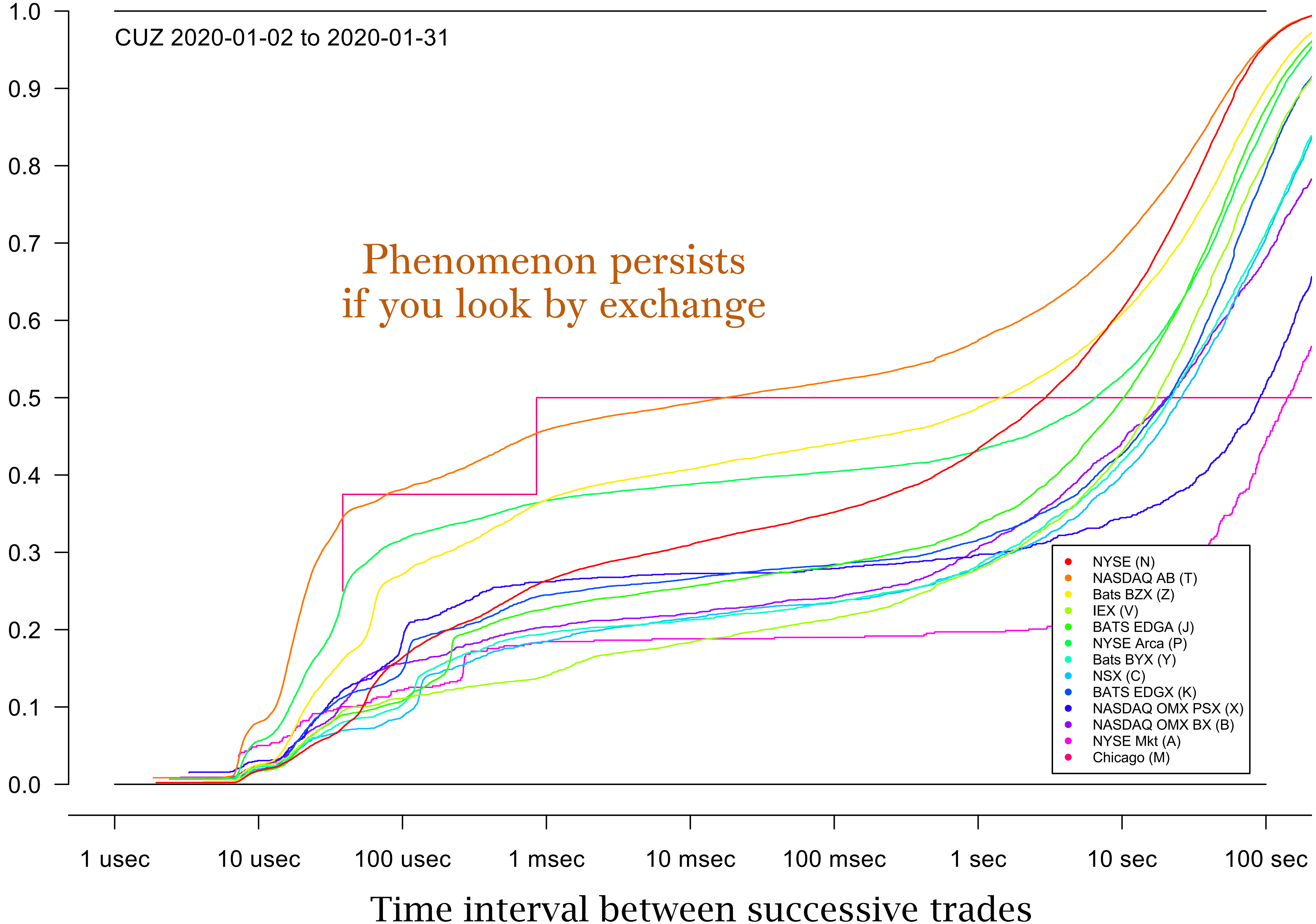
Can use "nanotime" package:
int64 number of nanoseconds since epoch

Time intervals between successive trades

```
T <- T[ order(T$date,T$time) ]
dtm = diff(T$time)
```




CUZ 2020-01-02 to 2020-01-31
Cumulative distribution
60 % of time differences are in this "small" category
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1 usec
10 usec
100 usec
1 msec
10 msec
100 msec
1 sec
10 sec
100 sec
Time interval between successive trades


CUZ 2020-01-02 to 2020-01-31
Phenomenon persists
if you look by exchange
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
NYSE (N)
NASDAQ AB (T)
Bats BZX (Z)
IEX (V)
BATS EDGA (J)
NYSE Arca (P)
Bats BYX (Y)
NSX (C)
BATS EDGX (K)
NASDAQ OMX PSX (X)
NASDAQ OMX BX (B)
NYSE Mkt (A)
Chicago (M)
1 usec
10 usec
100 usec
1 msec
10 msec
100 msec
1 sec
10 sec
100 sec
Time interval between successive trades

CODE	Description
A	NYSE MKT LLC
B	NASDAQ OMX BX, Inc.
C	National Stock Exchange Inc. (NSX)
D	Financial Industry Regulatory Authority, Inc. (FINRA ADF)
I	International Securities Exchange, LLC (ISE)
J	Bats EDGA Exchange, INC
K	Bats EDGX Exchange, Inc.
M	Chicago Stock Exchange, Inc. (CHX)
N	New York Stock Exchange LLC
P	NYSE Arca, Inc.
S	Consolidated Tape System
T	NASDAQ Stock Exchange, LLC (in Tape A, B securities)
Q	NASDAQ Stock Exchange, LLC (in Tape C securities)
V	The Investors' Exchange, LLC (IEX)
W	Chicago Broad Options Exchange, Inc.(CBOE)
X	NASDAQ OMX PSX, Inc. LLC
Y	Bats BYX Exchange, Inc.
Z	Bats BZX Exchange, Inc.

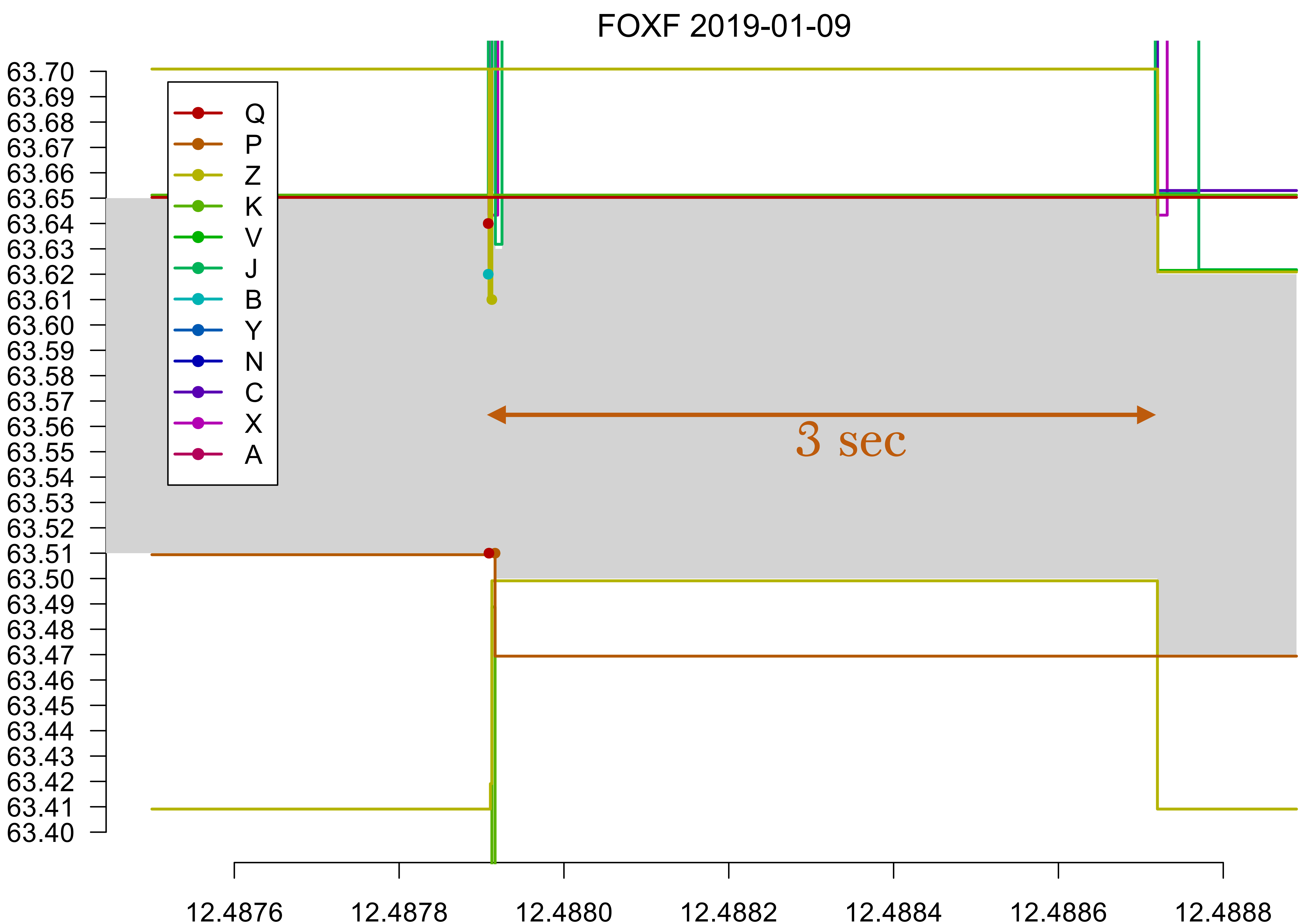
```
Q = NASDAQ
B = NASDAQ OMX
Z = Bats BZX
P = NYSE Arca
J = Bats EDGA
```

UTP issues

- @ = Regular Trade
- A = Acquisition
- B = Bunched Trade
- C = Cash Sale
- D = Distribution
- E = Placeholder for future use
- F = Intermarket Sweep
- G = Bunched Sold Trade
- H = Price Variation Trade
- I = Odd Lot Trade

15 sec before and after

```
> T[ time>44940 & time<44970 ]
         date          time exch  sym   prc siz cond corr     seq src rf                dt
1: 2019-01-09 44956.4692431    Q FOXF 63.64 100    @    0 1227975   N     6.53923911930e+01
2: 2019-01-09 44956.4692467    Q FOXF 63.64 100    @    0 1227976   N     3.59799742000e-06
3: 2019-01-09 44956.4692486    B FOXF 63.62 100    @    0 1227977   N     1.89000274986e-06
4: 2019-01-09 44956.4724100    Q FOXF 63.51   5 @F I    0 1227980   N     3.16134899913e-03
5: 2019-01-09 44956.4724882    Z FOXF 63.51   1 @F I    0 1227981   N     7.82349961810e-05
6: 2019-01-09 44956.4728166    P FOXF 63.51 100   @F    0 1227983   N     3.28378002450e-04
7: 2019-01-09 44956.4849461    Z FOXF 63.61 100    @    0 1227986   N     1.21295349963e-02
8: 2019-01-09 44956.4990845    P FOXF 63.51 100    @    0 1227989   N     1.41383810042e-02
```



FOXF 2019-01-09
63.70
63.69
63.68
63.67
63.66
63.65
63.64
63.63
63.62
63.61
63.60
63.59
63.58
63.57
63.56
63.55
63.54
63.53
63.52
63.51
63.50
63.49
63.48
63.47
63.46
63.45
63.44
63.43
63.42
63.41
63.40
Q
P
Z
K
V
J
B
Y
N
C
X
A
3 sec
12.4876
12.4878
12.4880
12.4882
12.4884
12.4886
12.4888

Quote data

Use R data.table and fasttime
for efficient loading of csv files

```
> head( Q[TIME_M>=10*60*60,], 20 )
          DATE         TIME_M EX   BID BIDSIZ   ASK ASKSIZ QU_COND QU_SEQNUM NATBBO_IND QU_CANCEL QU_SOURCE SYM_ROOT
 1: 2020-01-02 36000.0241680  J 40.76      3 40.80      1       R 202442501          A                   C      CUZ
 2: 2020-01-02 36000.0519836  Z 40.78      1 40.81      2       R 202449601          A                   C      CUZ
 3: 2020-01-02 36000.0520537  Z 40.78      1 40.80      1       R 202449701          A                   C      CUZ
 4: 2020-01-02 36000.1599755  J 40.76      3 40.83      3       R 202489201          A                   C      CUZ
 5: 2020-01-02 36000.1601002  Z 40.78      1 40.81      1       R 202489301          A                   C      CUZ
 6: 2020-01-02 36001.0543792  K 40.13      2 40.95      1       R 202739701          A                   C      CUZ
 7: 2020-01-02 36001.1555188  K 40.13      2 40.95      1       R 202756901          A                   C      CUZ
 8: 2020-01-02 36003.0105681  N 40.78      6 40.80      1       R 203009501          U                   C      CUZ
 9: 2020-01-02 36003.0105774  N 40.78      5 40.80      1       R 203009601          U                   C      CUZ
10: 2020-01-02 36003.0105910  Z 40.78      1 40.81      1       R 203009701          A                   C      CUZ
11: 2020-01-02 36003.0106161  N 40.78      4 40.80      1       R 203009801          U                   C      CUZ
12: 2020-01-02 36003.0106246  N 40.78      3 40.80      1       R 203009901
13: 2020-01-02 36003.0107245  Y 40.64      1 40.94      1       R 203010001
14: 2020-01-02 36003.0110040  X 40.78      5 40.91      1       R 203010201
15: 2020-01-02 36003.0110230  X 40.78      6 40.91      1       R 203010301
16: 2020-01-02 36004.0581398  K 40.13      2 40.95      1       R 203122302
17: 2020-01-02 36004.0585508  X 40.78      6 40.80      5       R 203122601
18: 2020-01-02 36004.0585647  X 40.78      6 40.80      6       R 203122701
19: 2020-01-02 36004.0763361  K 40.13      2 40.95      1       R 203122801
20: 2020-01-02 36004.5376861  K 40.13      2 40.95      1       R 203207901
```

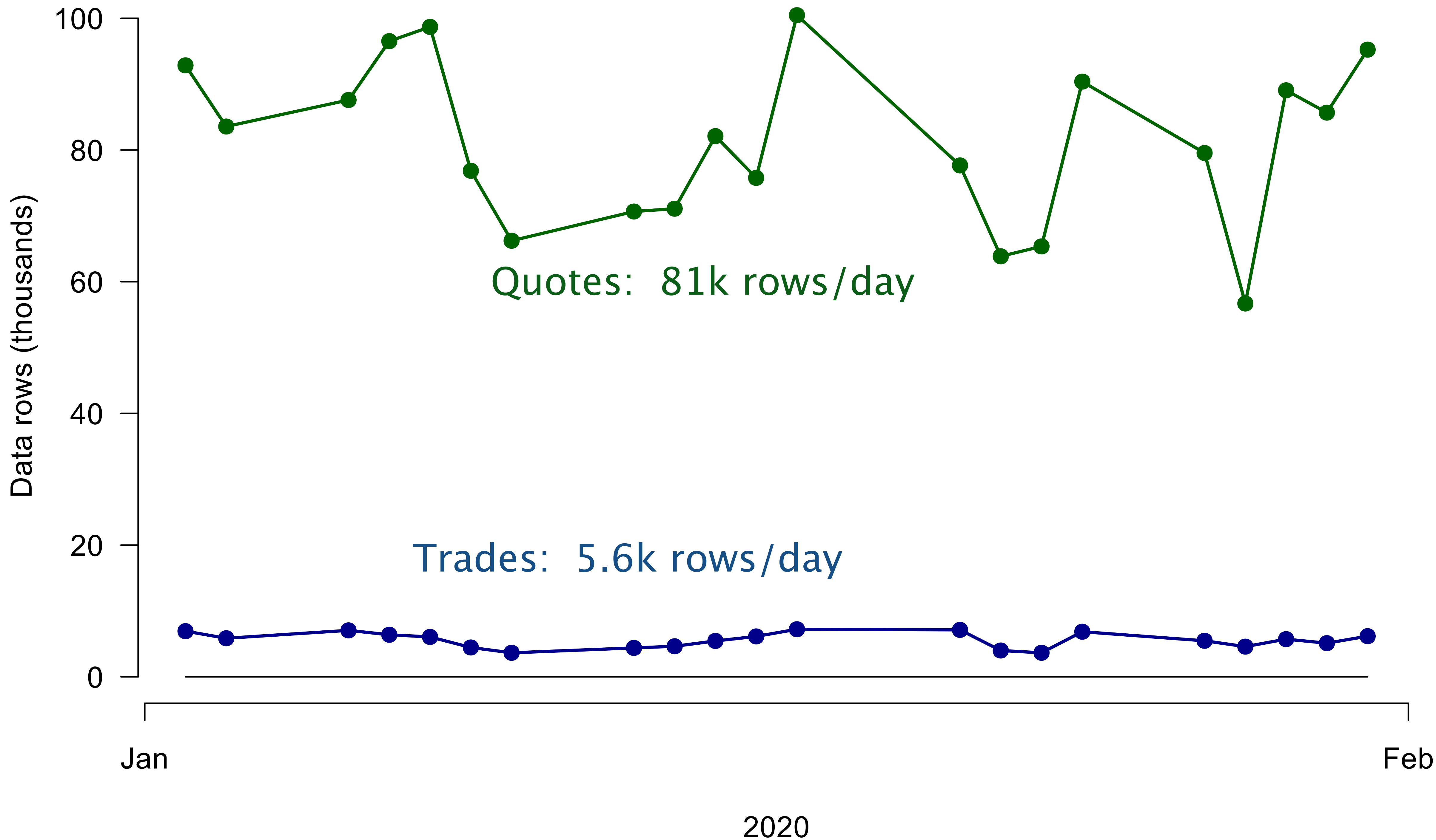
Quotes

	ex	n
1	T	296486
2	N	245209
3	Z	211876
4	P	189908
5	J	172384
6	Y	160850
7	C	145541
8	K	122103
9	X	69212
10	B	42782
11	A	39303
12	V	7119
13	M	3072

Trades

	ex	n
1	T	28691
2	N	23883
3	Z	15404
4	J	11383
5	P	10764
6	V	7214
7	K	6767
8	C	4143
9	Y	4128
10	B	2526
11	X	1402
12	A	559
13	M	9

CUZ Jan 2020



100
80
60
40
20
0
Data rows (thousands)
Quotes: 81k rows/day
Trades: 5.6k rows/day
Jan
Feb
2020

THE WALL STREET JOURNAL.

U.S. Edition ▾ | February 16, 2019 | Print Edition | Video

MARKETS

Brief Price Gaps in Stocks Cost Investors $2 Billion a Year

Government-backed study finds nearly 24% of trades may not be executed at best available prices, can be exploited by fast traders

By *Cezary Podkul*

25 COMMENTS

Feb. 14, 2019 9:00 a.m. ET

—Alexander Osipovich contributed to this article.

Nearly a quarter of U.S. equity trades may not be executed at the best price available in the market, costing investors at least $2 billion a year, according to a new study funded by the U.S. government.

They found that the national best bid or offer often wasn't really the best quote available in the market at the time of the trade. One trade completed at 9:48:55.398272 sold 100 shares of Apple for $99.12—then the national best bid available in the market—but direct feeds from exchanges showed 100 shares could have been sold elsewhere for $99.16, costing the seller $4 on the $9,912 transaction.

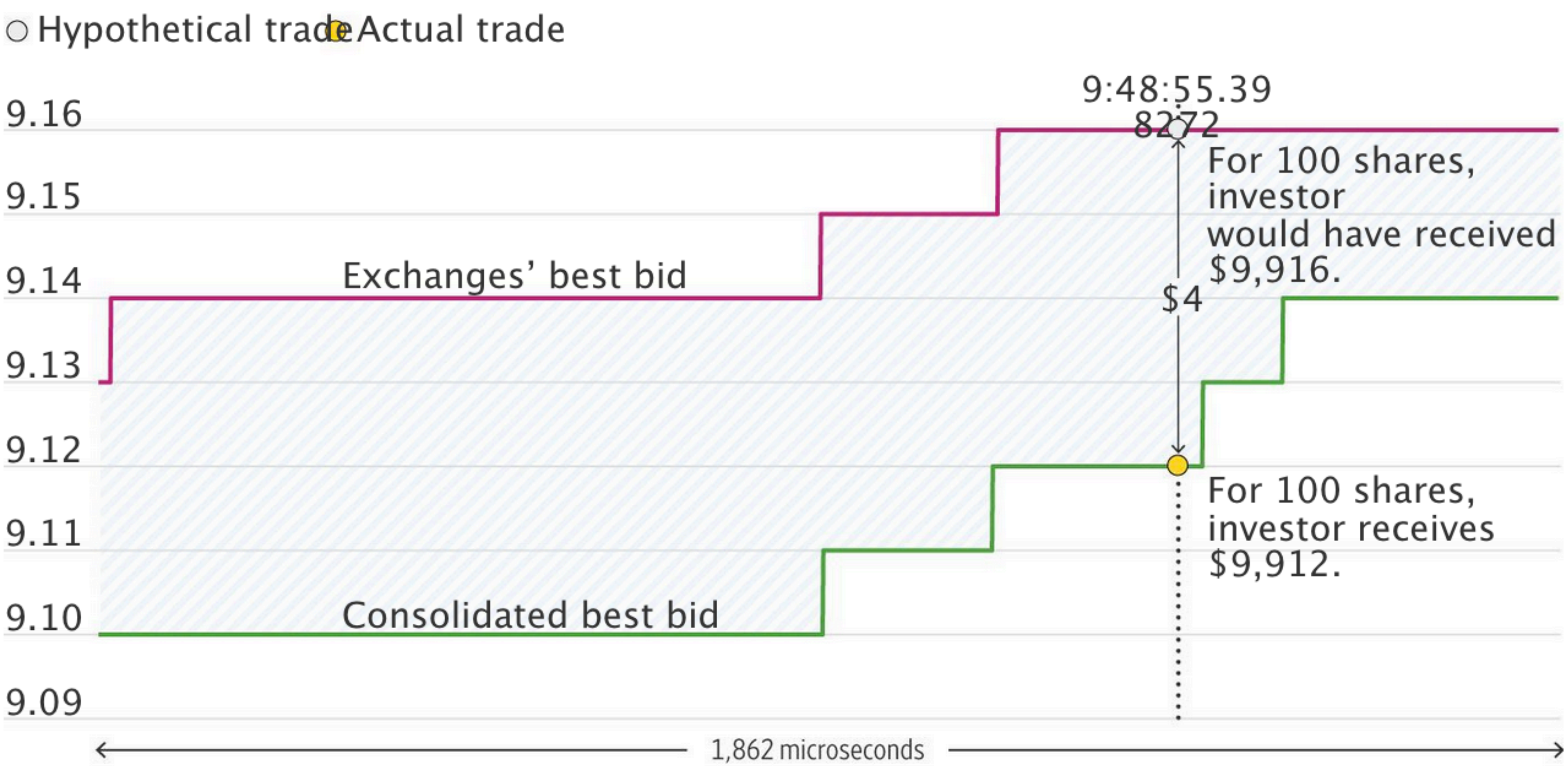
Altogether, Mr. Tivnan's team calculated that investors buying and selling Apple stock on Jan. 7, 2016, may have been able to save themselves at least $105,000 by trading at prices published on the exchanges' direct data feeds. Across all of 2016, the researchers tabulated a total opportunity cost of about $2 billion, which they say is a conservative estimate.

https://www.wsj.com/articles/brief-price-gaps-in-stocks-cost-investors-2-billion-a-year-11550152800

In a Flash

There is usually just one best bid and one best offer price for stocks. But researchers spotted brief gaps between the best available price nationally and the best price on specific exchanges, a divergence that gives high-speed traders a chance to profit.

Apple share price around Jan. 7, 2016, at 9:48:55 a.m. ET



Note: Consolidated bid refers to national best bid observed in Securities Information Processor data. Exchanges' best bid refers to best bid observed in direct data feeds from stock exchanges.

Sources: Data from Thesys Technologies LLC, analysis by MITRE Corp. and University of Vermont.

https://www.wsj.com/articles/brief-price-gaps-in-stocks-cost-investors-2-billion-a-year-11550152800

RESEARCH ARTICLE

Fragmentation and inefficiencies in US equity markets: Evidence from the Dow 30

Brian F. Tivnan[1,2,3,4]*, David Rushing Dewhurst[2,4,5,6], Colin M. Van Oort[1,2,4,5,6], John H. Ring IV[1,2,4,5,6], Tyler J. Gray[2,3,4,5], Brendan F. Tivnan[4,7], Matthew T. K. Koehler[1,4], Matthew T. McMahon[1,4], David M. Slater[1], Jason G. Veneman[1,4], Christopher M. Danforth[2,3,4,5]*

1 The MITRE Corporation, McLean, VA, United States of America, **2** Vermont Complex Systems Center, University of Vermont, Burlington, VT, United States of America, **3** Department of Mathematics and Statistics, University of Vermont, Burlington, VT, United States of America, **4** Computational Finance Lab, Burlington, VT, United States of America, **5** Department of Computer Science, University of Vermont, Burlington, VT, United States of America, **6** Computational Story Lab, University of Vermont, Burlington, VT, United States of America, **7** School of Engineering, Tufts University, Medford, MA, United States of America

* btivnan@mitre.org (BFT); chris.danforth@uvm.edu (CMD)

PLOS ONE | https://doi.org/10.1371/journal.pone.0226968 January 22, 2020

Using the most comprehensive source of commercially available data on the US National Market System, we analyze all quotes and trades associated with Dow 30 stocks in calendar year 2016 from the vantage point of a single and fixed frame of reference. We find that inefficiencies created in part by the fragmentation of the equity marketplace are relatively common and persist for longer than what physical constraints may suggest. Information feeds reported different prices for the same equity more than 120 million times, with almost 64 million dislocation segments featuring meaningfully longer duration and higher magnitude. During this period, roughly 22% of all trades occurred while the SIP and aggregated direct feeds were dislocated. The current market configuration resulted in a realized opportunity cost totaling over $160 million, a conservative estimate that does not take into account intra-day offsetting events.

https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0226968

We study the behavior of these securities as traded in modern US equity markets, known as the National Market System (NMS). The NMS is comprised of 13 networked exchanges coupled by information feeds of differential quality and subordinated to national regulation. Adding another layer of complexity, the NMS supports a diverse ecosystem of market participants, ranging from small retail investors to institutional financial firms and designated market makers.

We do not attempt to unravel and attribute the activity of each of these actors here; several others have attempted to classify such activities with varying degrees of success in diverse markets [2–4]. We take a first-principles approach by compiling an exhaustive catalog of every dislocation, defined as a nonzero pairwise difference between the prices displayed by the National Best Bid and Offer (NBBO), as observed via the Securities Information Processor (SIP) feed, and Direct Best Bid and Offer (DBBO), as observed via the consolidation of all direct feeds.

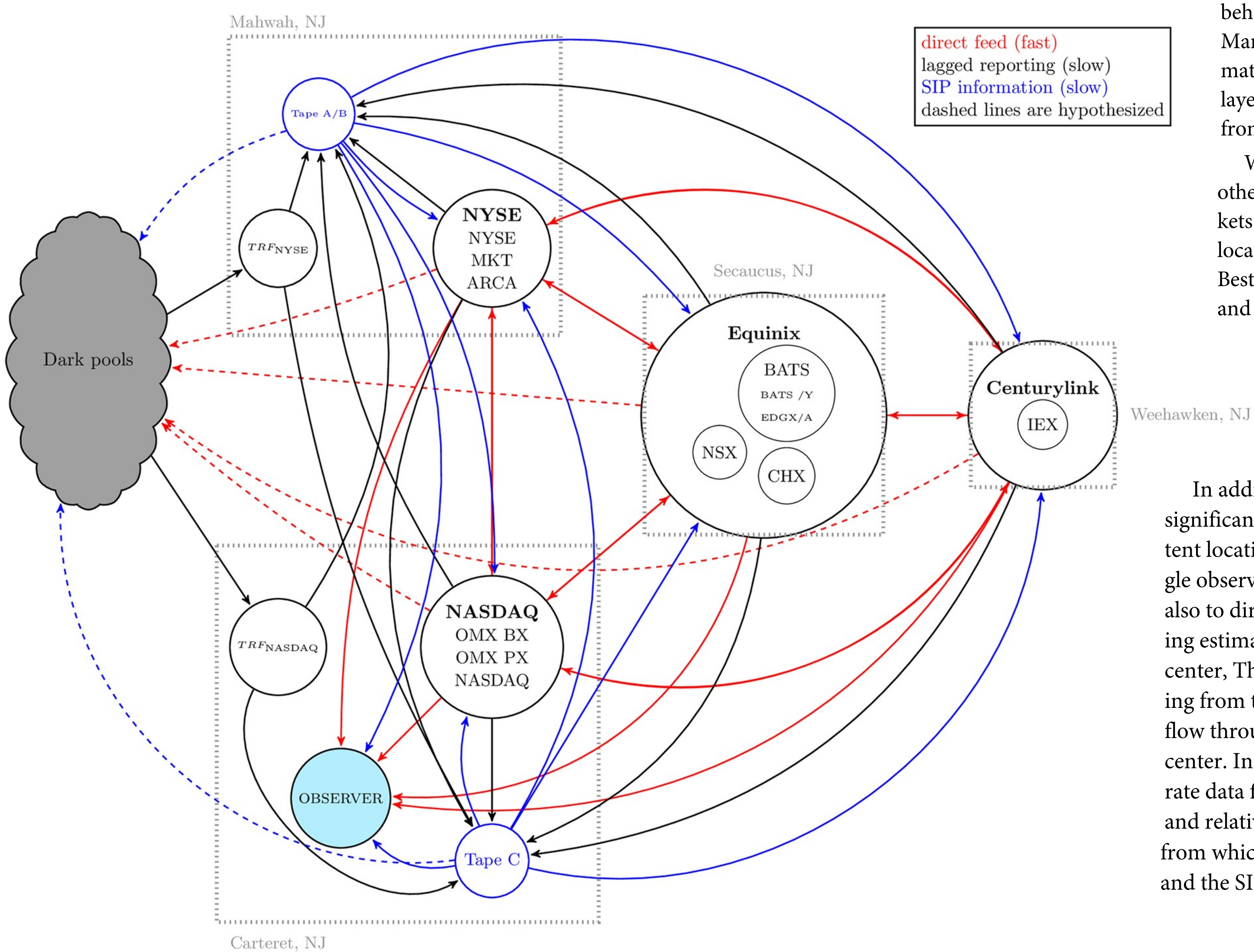


Fig 1. The NMS (lit market and ATSs) as implied by the comprehensive market data. As we do not have the specifications of inter-market center communication mechanisms and have minimal knowledge of intra-market center communication mechanisms, we simply classify information as having high latency, as the SIP and lagged information heading to the SIP do, or low latency, as the information on the direct feeds does. Note the existence of the observer, located in Carteret NJ. Without a single, fixed observer it is difficult to clock synchronization issues and introduces an unknown amount of noise into measurements of dislocations and similar phenomena. Clock synchronization issues are avoided when using data collected from a single point of presence since all messages may be timestamped by a single clock, controlled by the observer.

https://doi.org/10.1371/journal.pone.0226968.g001

In addition to being the authoritative data source for the SEC's MIDAS program, another significant attribute of the Thesys data is that it is collected by a single observer from a consistent location in the NMS (the Nasdaq data center in Carteret, NJ) as depicted in Fig 1. The single observer not only allows the user to account for the relativistic effects described above but also to directly observe dislocation segments and realized opportunity cost instead of compiling estimates of these quantities as has been done in previous studies. At the NASDAQ data center, Thesys applies a new timestamp to each message received, including messages originating from the SIP feed or one of the direct feeds, that allows subscribers to observe information flow through the NMS in the same manner as a market participant located at the Carteret data center. In our analysis we use this "Thesys timestamp" to synchronize information from disparate data feeds and avoid issues that otherwise could arise from clock synchronization errors and relativistic effects. Since this timestamp is given at the time the data arrives at the server from which the data is collected, any discrepancies in the clocks at different exchanges, ATSs, and the SIP do not affect our measurement procedures.

MODERN MARKETS INITIATIVE

February 14, 2019

Modern Markets Initiative (MMI) is the education and advocacy organization devoted to the role of technological innovation in creating the world's best markets. It was founded in 2013 by leading, quantitative trading firms Global Trading Systems, Hudson River Trading, Quantlab Financial and Tower Research Capital whose collective role in creating market efficiency contributes to a dramatic reduction in trading costs for investors. MMI engages and educates public audiences about the value modern market professionals provide to today's electronic marketplace. Additional information is available at www.modernmarkets.org.

Mitre / University of Vermont Study: Analysis from MMI Industry Peer Review

Modern Markets Initiative ("MMI"), the education and advocacy association devoted to the role of technological innovation in creating the world's best markets, is greatly appreciative of the opportunity to a provide an internal industry review and factual analysis of the Mitre /University of Vermont study on market structure (collectively, "the Mitre Study").

(1) SIP CLASSIFICATION AS A "TRADING VENUE" IS A FALSE ASSUMPTION

MMI's peer review group noted that the paper refers to entities that "buy on the SIP." The author's hypothesis that you can buy on the SIP seems to be the foundation of the paper's claims about latency arbitrage. However, MMI notes that it is impossible to buy on the SIP because the SIP is not a trading venue, but rather a reporting venue of prices (both trades and quotes). Similarly, the direct feeds offered by exchanges only report trades and quotes. You cannot "buy" on these feeds either. The authors seem to confuse trade and quote reporting mechanisms with actual exchanges. When you send an order to an exchange, you will receive the price that is available on the exchange at that time, regardless of what the SIP or the direct feeds may have indicated when you decided to send the order. A discrepancy between the SIP and the direct data feed is thus not an opportunity to make profits from an arbitrage just as you cannot make profits from "trading" on the difference between a price published in *The Wall Street Journal* versus a price published in the *Financial Times*. Further, the authors seem to mistakenly think that the SIP and direct feeds are information streams from two different markets, when in fact they are streams about the same market arriving at two slightly different times.

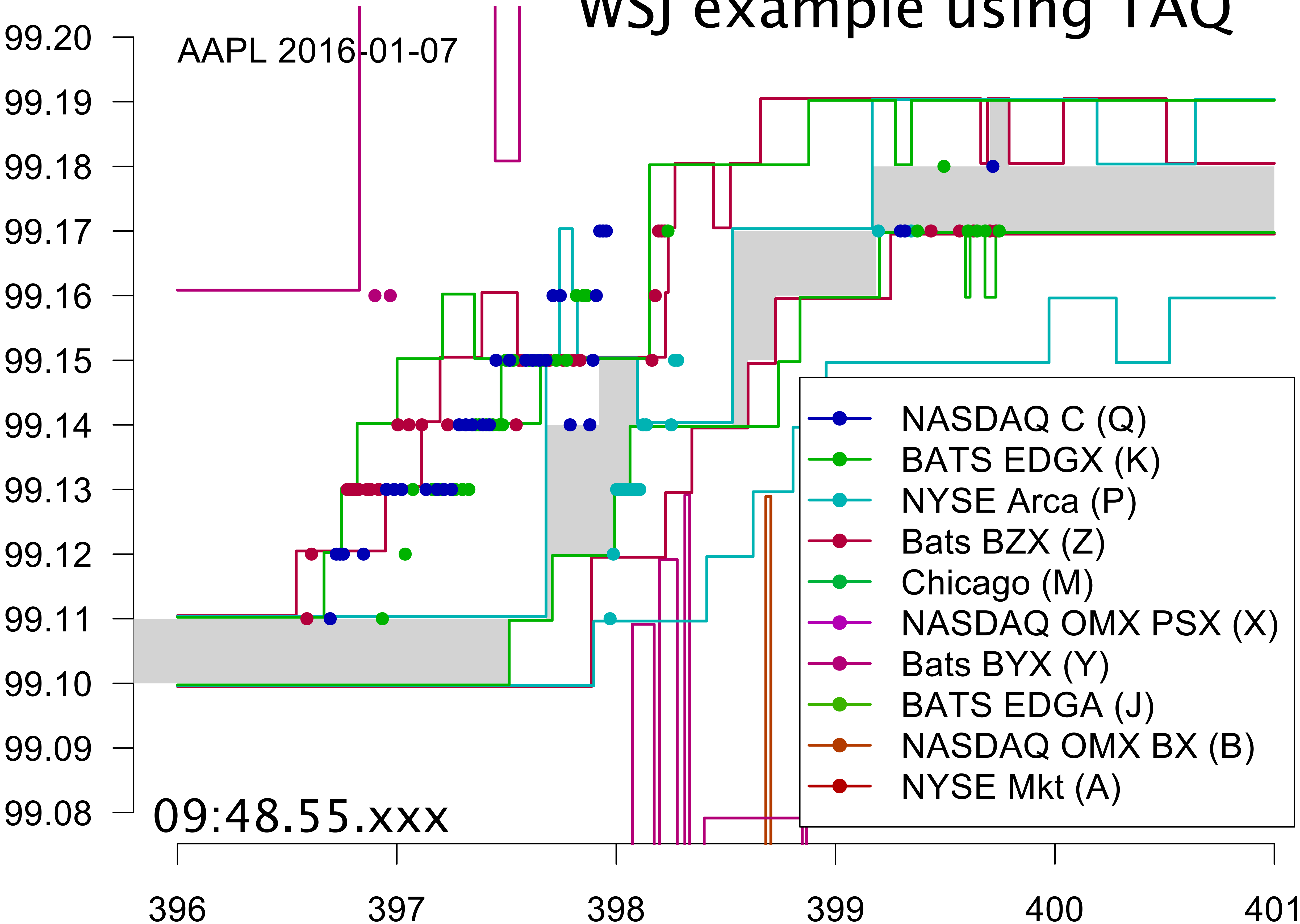
(3) FAILURE TO ACCOUNT FOR REGULATION NMS

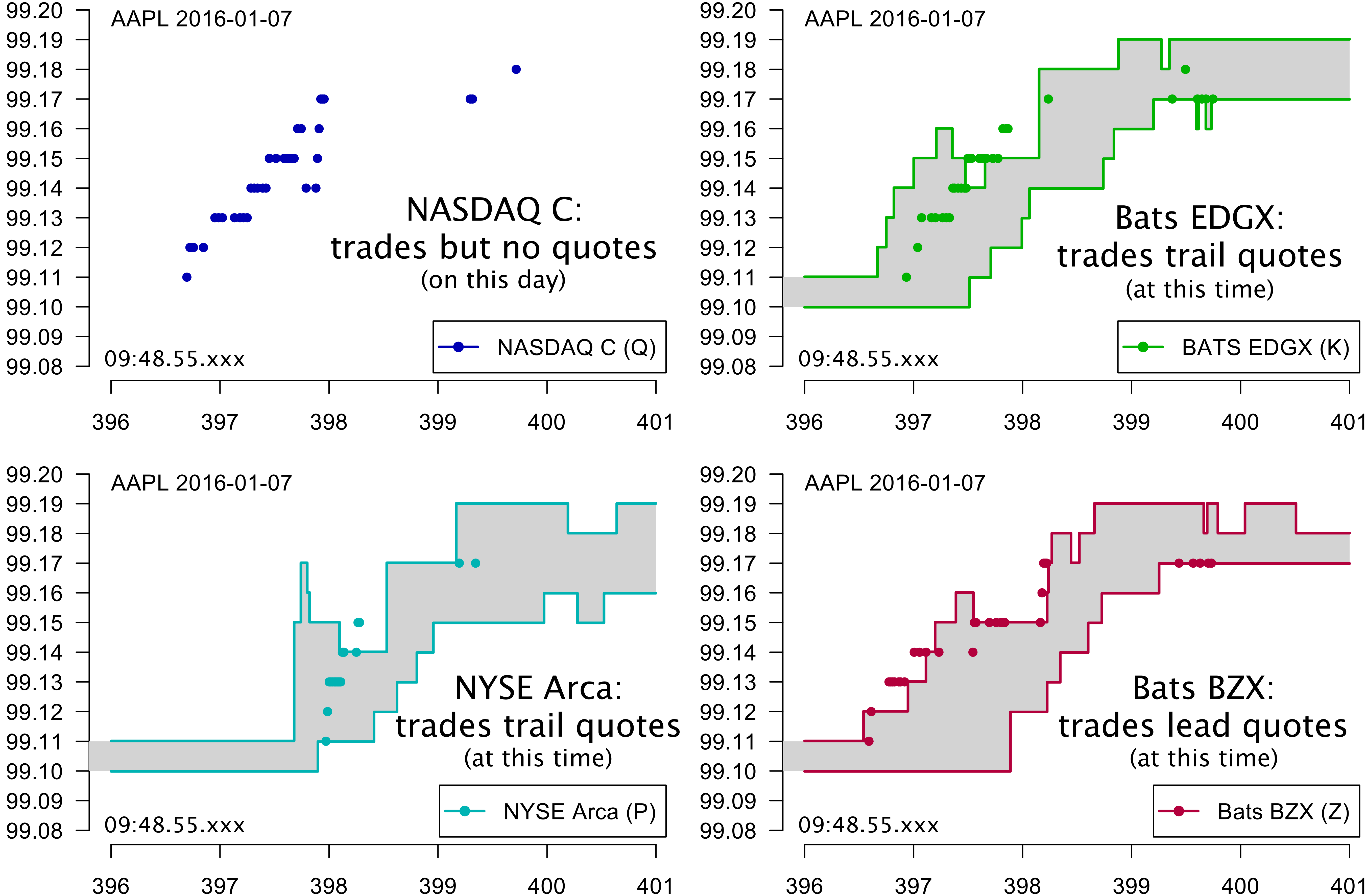
In addition to the authors' incorrect interpretation of what the SIP is, and what it is not, the authors also provide an example of latency arbitrage that fails to account for core market structure rules embedded in Regulation NMS. In their example the authors postulate the following hypothetical they believe creates an opportunity for someone to profit from latency arbitrage:

1. *Suppose there are two exchanges, A and B, which trade a single security, S.*
2. *Assume that a trader is located at A and that the BBO for S is initially bid @ $100.00 - ask @ $100.01 at both A and B.*
3. *At some future time, the BBO at A changes to bid @ $99.98 - ask @ $99.99 while the BBO at B stays the same.*
4. *The trader purchases shares of S at $99.99 from A and sells shares at B for $100.00, obtaining a gross profit of $0.01 per share.*

Step 3 above is known as a crossed-market. It is not allowed under Reg NMS and exchanges must have policies and procedures to prevent such occurrences. Though it is certainly possible that the bid at exchange A changes to $99.98 in step 2, exchange A is prohibited from posting an offer at $99.99 while there is a bid for $100.00 on exchange B. In practice exchange A would need to route the order to sell at $99.99 to exchange B where it would hit the bid on exchange B and be filled at $100.00. In general, you don't get to step 4, so the authors' hypothetical trader never gets the opportunity they suggest.

WSJ example using TAQ





AAPL 2016-01-07
NASDAQ C:
trades but no quotes
(on this day)
NASDAQ C (Q)
09:48.55.xxx
AAPL 2016-01-07
Bats EDGX:
trades trail quotes
(at this time)
BATS EDGX (K)
09:48.55.xxx
AAPL 2016-01-07
NYSE Arca:
trades trail quotes
(at this time)
NYSE Arca (P)
09:48.55.xxx
AAPL 2016-01-07
Bats BZX:
trades lead quotes
(at this time)
Bats BZX (Z)
09:48.55.xxx

Matching trades and quotes

- Necessary to evaluate trades to bid/ask
- Necessary to properly determine trade direction
- Possible only approximately with TAQ (SIP)
- Sometimes possible with direct exchange feeds

Trade sign determination from quotes

- match trade price to prevailing NBBO bid/ask
  allowing for roundoff error

  prc > mid + $\varepsilon \Longrightarrow$ buy

  prc < mid - $\varepsilon \Longrightarrow$ sell

  else $\Longrightarrow$ "tick test" or ignore
- Use time stamp since TAQ sequence numbers
  are not consistent across trade/quote and exchanges
  (though time stamps are imperfect)
- this is sign of *aggressive* order
  maybe not sign of your order

Roundoff error

ε smaller than tick, say 0.001

Trade price

prc > ask

Ask — prc = ask

bid < prc < ask

Bid — prc = bid

prc < bid

Trade price

prc > ask + **ε**

Ask — ask-**ε** ≤ prc ≤ ask+**ε**

bid+ **ε** < prc < ask-**ε**

Bid — bid-**ε** ≤ prc ≤ bid+**ε**

prc < bid-**ε**

Example: 2018-01-25

Few trades outside NBBO spread (timing errors?)

CTLT

- nL = 20
- nB = 1614
- nM = 1886
- nA = 1805
- nH = 55

Many trades inside NBBO spread (hidden liquidity)

MSFT

- nL = 184
- nB = 16398
- nM = 7016
- nA = 16621
- nH = 459

Include only exchanges N, T, P, Z

Bid-ask spread modeling

- Significance
  - Measure of liquidity: spread is cost of small trade
- Empirical
  - Measuring the spread and its average value
  - Roll model
  - Trade sign identification (Lee-Ready algorithm)
  - Statistical properties of trade sign: serial correlation
- Theoretical models: why is there a spread?
  - Fixed costs (Roll)
  - Information (Glosten-Milgrom)
  - Inventory risk (Stoll)

NYSE

ICE NYSE ABOUT CONTACT

LOGIN LIST TRADE DATA GOVERNANCE SERVICES INSIGHTS REGULATION

Spread and size are indicators of market quality

MARKETS

NYSE

With almost 3 times more liquidity than the next largest exchange, the most displayed shares at the NBBO, and the narrowest quoted bid / ask spreads, NYSE is the global leader in market quality. As the world's markets evolve, NYSE responds with leading technology, innovative partnerships and trading techniques.

https://www.nyse.com/markets/nyse

Equity Trading in the 21st Century: An Update

James J. Angel
Associate Professor
McDonough School of Business
Georgetown University
angelj@georgetown.edu

Lawrence E. Harris
Fred V. Keenan Chair in Finance
Professor of Finance and Business Economics
Marshall School of Business
University of Southern California
LHarris@usc.edu

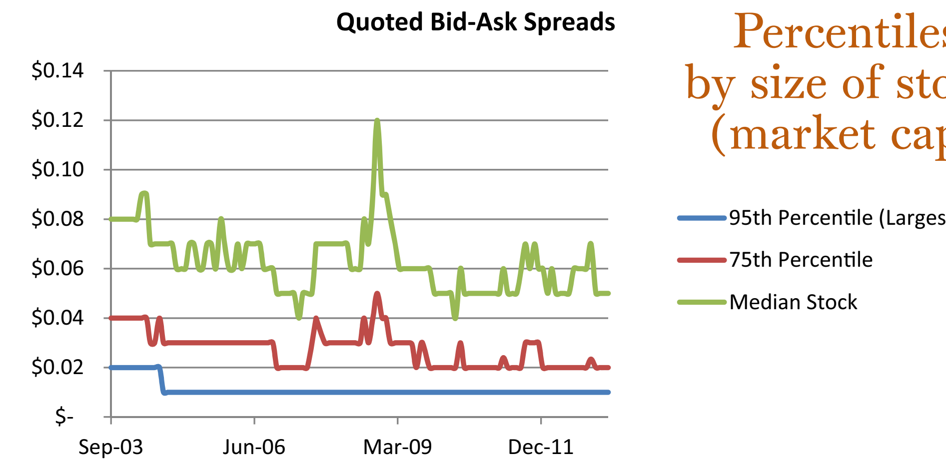
Chester S. Spatt
Pamela R. and Kenneth B. Dunn Professor of Finance
Tepper School of Business
Carnegie Mellon University
cspatt@andrew.cmu.edu

Published 20 January 2015

This paper updates our previous study, “Equity Trading in the 21st Century”, which presented results about US equity market quality. Despite many complaints in the national media, various measures of market quality indicate that US markets continue to be very healthy. Trade transaction cost estimates have stayed low and market depth and execution speeds remain high. New findings that measure the total transaction cost of executing very large block orders indicate that improvements in market quality also have benefited large institutional traders. While still high, both the number of quotes per trade and per minute have declined substantially from their peaks in 2008. Intraday volatility is below the levels of the pre-electronic 1990s. Although market quality is quite good, it could be enhanced. We discuss some current concerns about maker/taker pricing, dark pools, high frequency trading, tick sizes, designated dealers, transaction taxes, IPOs, and market stability.

1.1. *Summary of major results*

In general, most measures of market quality have remained fairly stable and at high levels over the last three years. Quotation activity has fallen sharply from its recent peaks, but remains surprisingly high to those who are unaware of how electronic traders operate. Overall, our results show that electronic trading has substantially improved markets and that investors have continued to enjoy those improvements over the last three years since we first documented these results.



Percentiles by size of stock (market cap)

Fig. 2. (Color online) Quoted bid-ask spreads.

2.2. *Bid-ask spreads remain small*

Figure 2 displays the quoted bid-ask spreads for securities traded in the US National Market System from 2003 through 2012. The median quoted bid-ask spread was calculated for each stock and percentiles of the cross-sectional distribution of these median quoted spreads are displayed here. The quoted spreads for the largest stocks remain at one cent. For smaller stocks, spreads jumped during the financial crisis, but since have returned to historic lows.

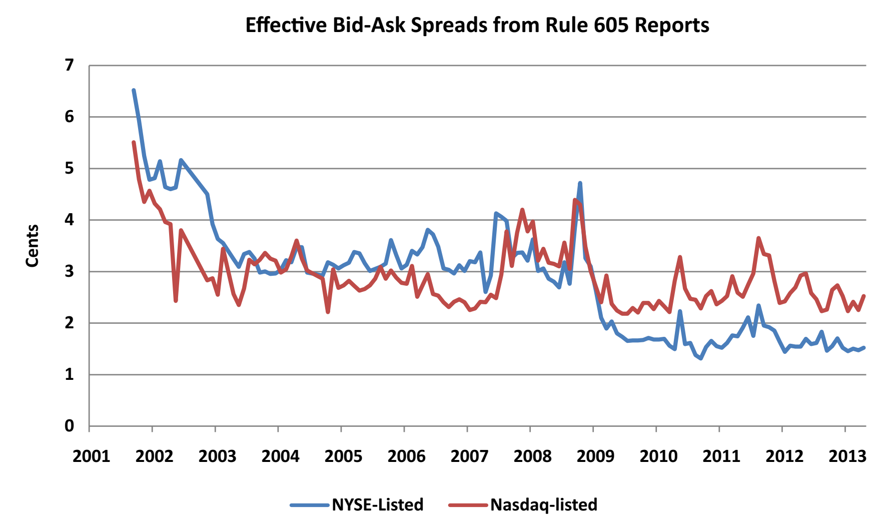


Fig. 4. (Color online) Effective bid-ask spreads from Rule 605 reports.

2.4. *Effective spreads also remain small*

Figure 4 displays average effective bid-ask spreads obtained from Rule 605 reports for eligible market orders. The effective bid-ask spread estimates spreads that investors actually pay. It is twice the difference between the actual trade price and the midpoint of the quoted National Best Bid or Offer (“NBBO”) at the time of order receipt. The general downward trend in spreads was interrupted by an upward spike during the recent financial crisis. Care should be used in interpreting differences across exchanges because the stocks that the two exchanges list differ in capitalizations, price levels, volumes, and volatilities — variables that all affect bid-ask spreads.

Spread (empirical)

- Exchange i has bid $B_i(t)$, ask $A_i(t)$

  piecewise constant functions of time

  in market data, will almost always have $A_i(t)>B_i(t)$
  (except where bid=0 or ask=0, or other glitches)

- NBBO Best bid $B(t)=\max_i B_i(t)$, best ask $A(t)=\min_i A_i(t)$

  Reg NMS says should have $A(t)>B(t)$
  but may not in data because of timing errors

  max,min may be taken over "relevant" exchanges

- NBBO Midpoint $M(t) = (B(t) + A(t)) / 2$

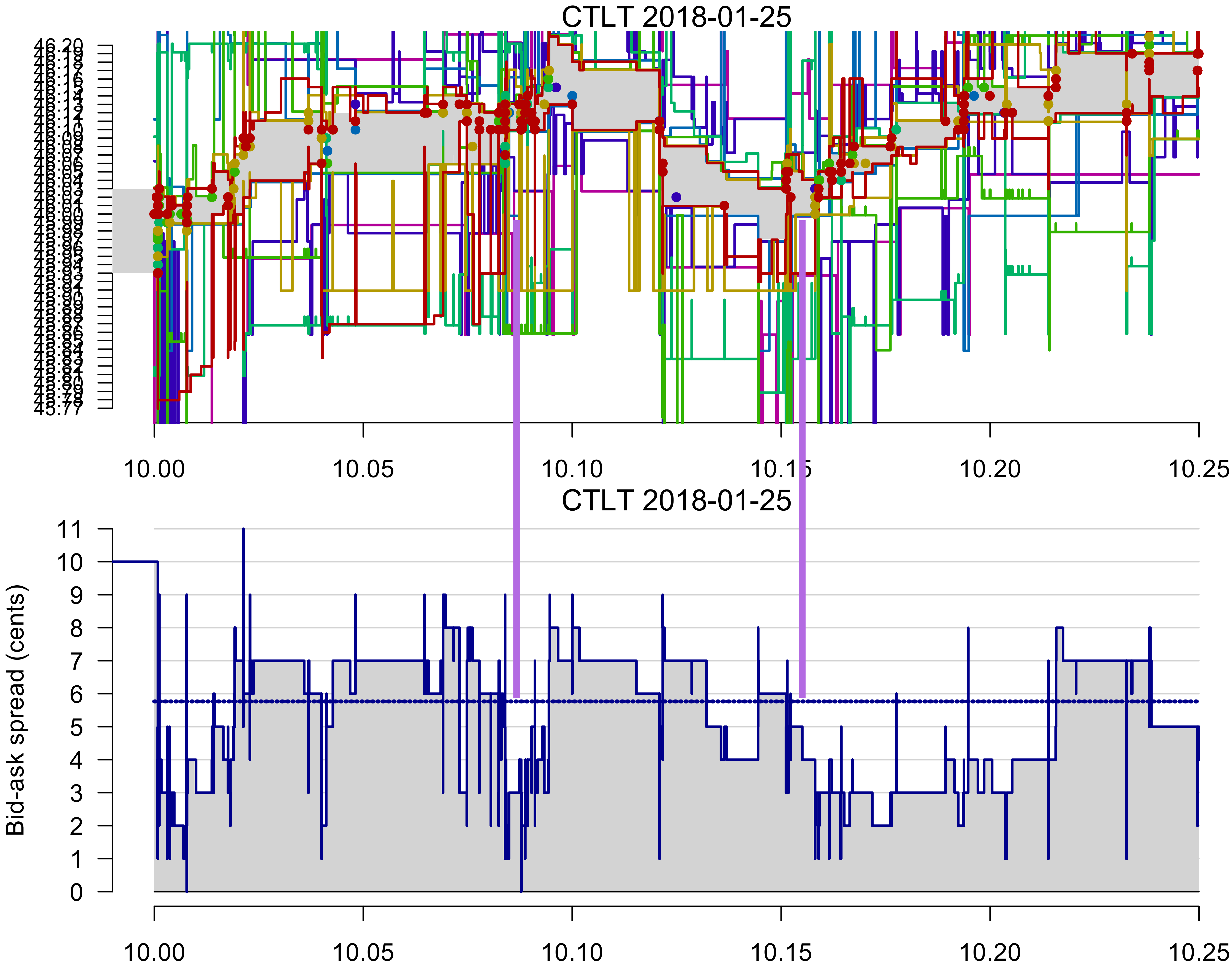
  piecewise constant function of time

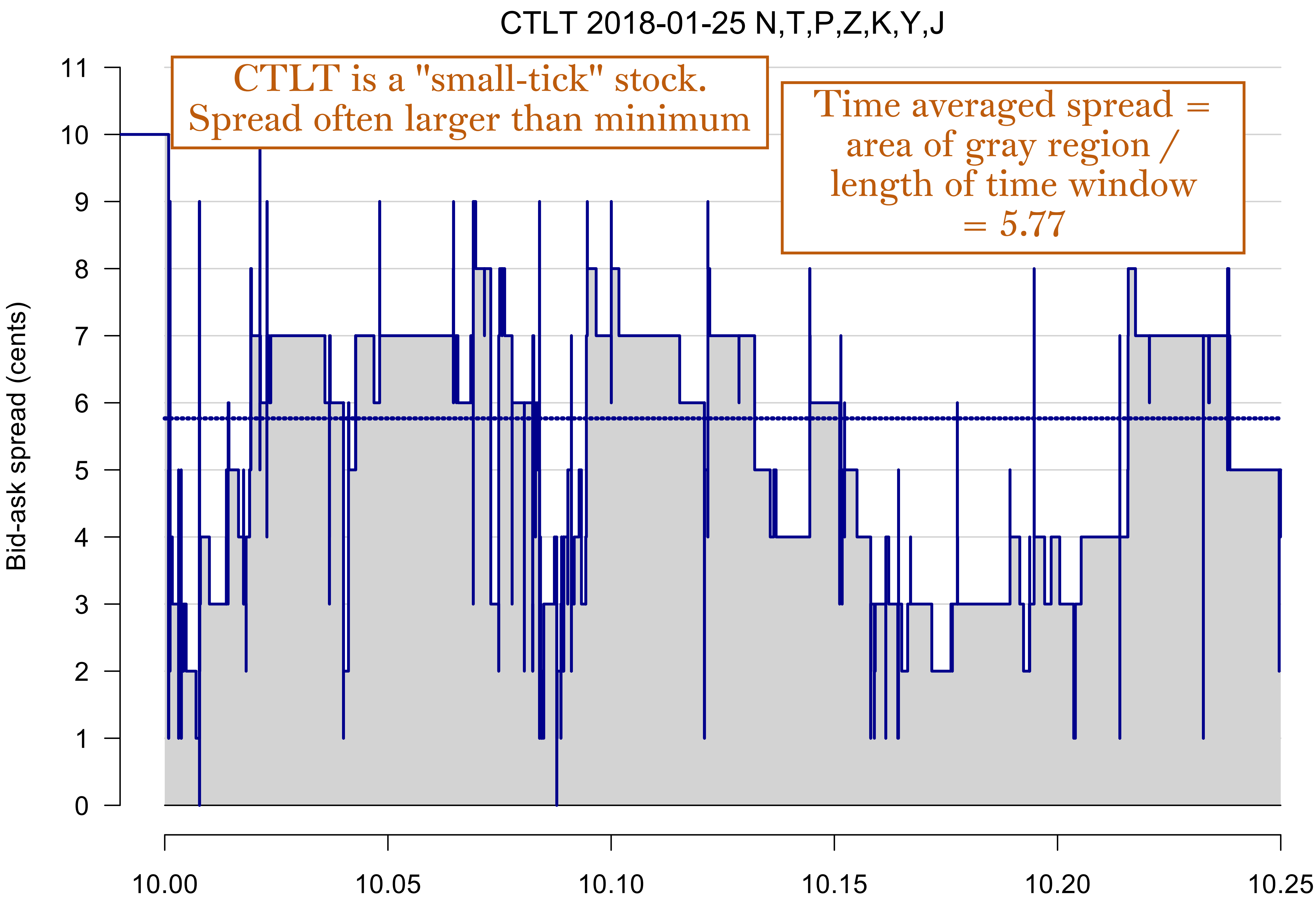
  use as the "price" when need unique price (volatility)

- NBBO Spread $S(t) = A(t) - B(t)$

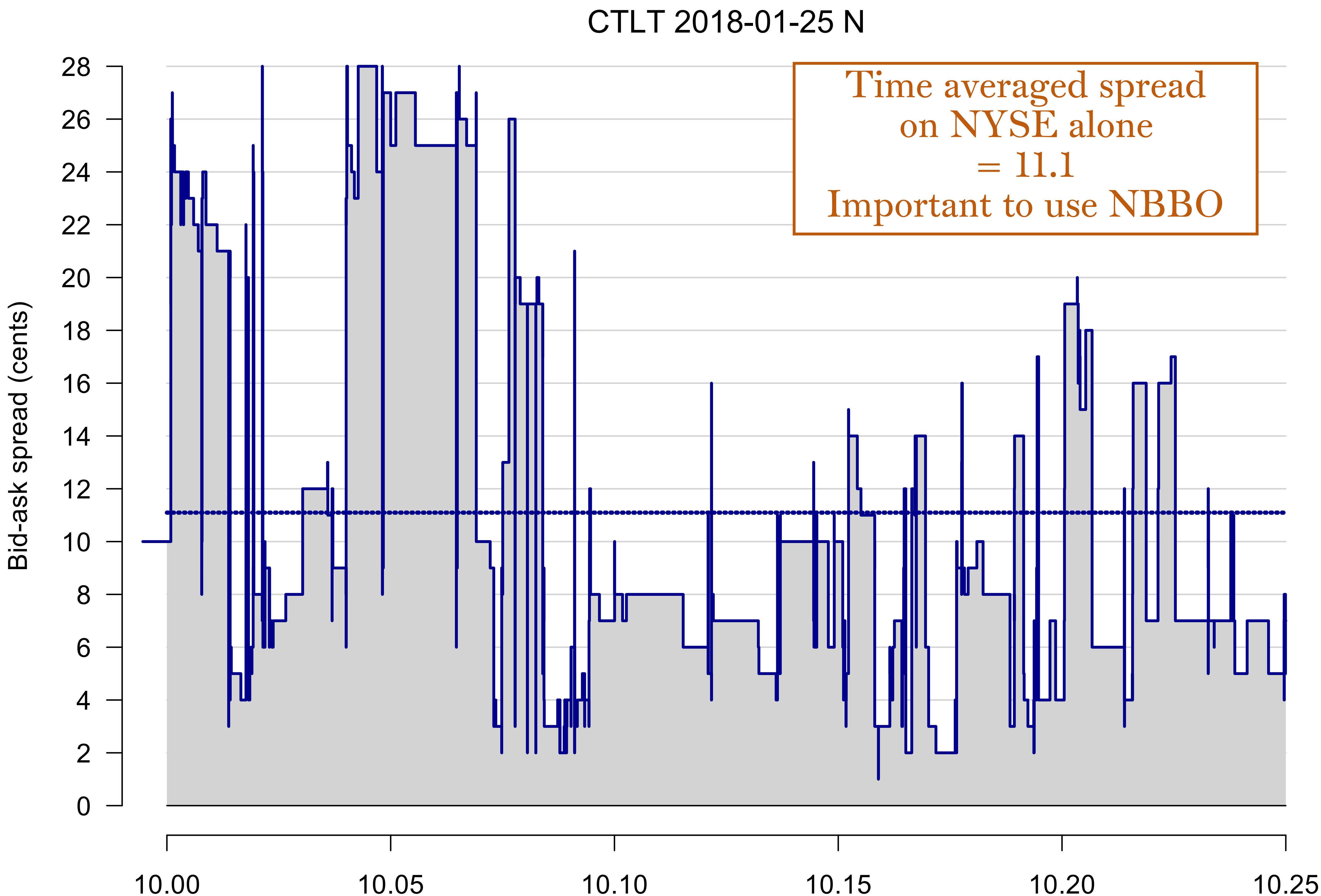
  compute averages in one of two ways:

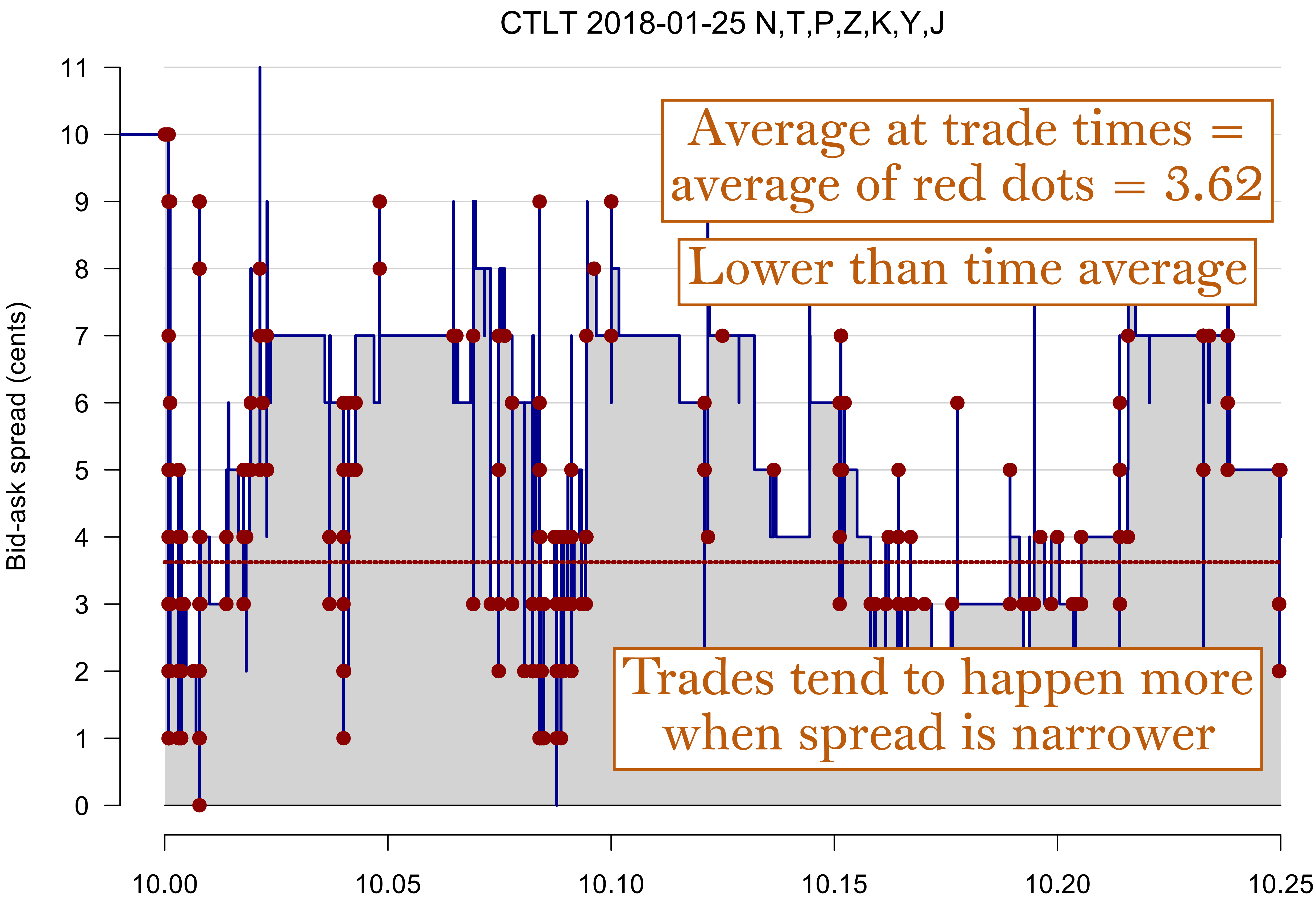
  - weighting over all time
  - weighting at trade times


CTLT 2018-01-25
10.00
10.05
10.10
10.15
10.20
10.25
CTLT 2018-01-25
Bid-ask spread (cents)
11
10
9
8
7
6
5
4
3
2
1
0
10.00
10.05
10.10
10.15
10.20
10.25


CTLT 2018-01-25 N,T,P,Z,K,Y,J
CTLT is a "small-tick" stock.
Spread often larger than minimum
Time averaged spread =
area of gray region /
length of time window
= 5.77
Bid-ask spread (cents)
0
1
2
3
4
5
6
7
8
9
10
11
10.00
10.05
10.10
10.15
10.20
10.25

If look only at one exchange




CTLT 2018-01-25 N,T,P,Z,K,Y,J
Bid-ask spread (cents)
0
1
2
3
4
5
6
7
8
9
10
11
10.00
10.05
10.10
10.15
10.20
10.25
Average at trade times = average of red dots = 3.62
Lower than time average
Trades tend to happen more when spread is narrower

Subtleties in averaging at trade times

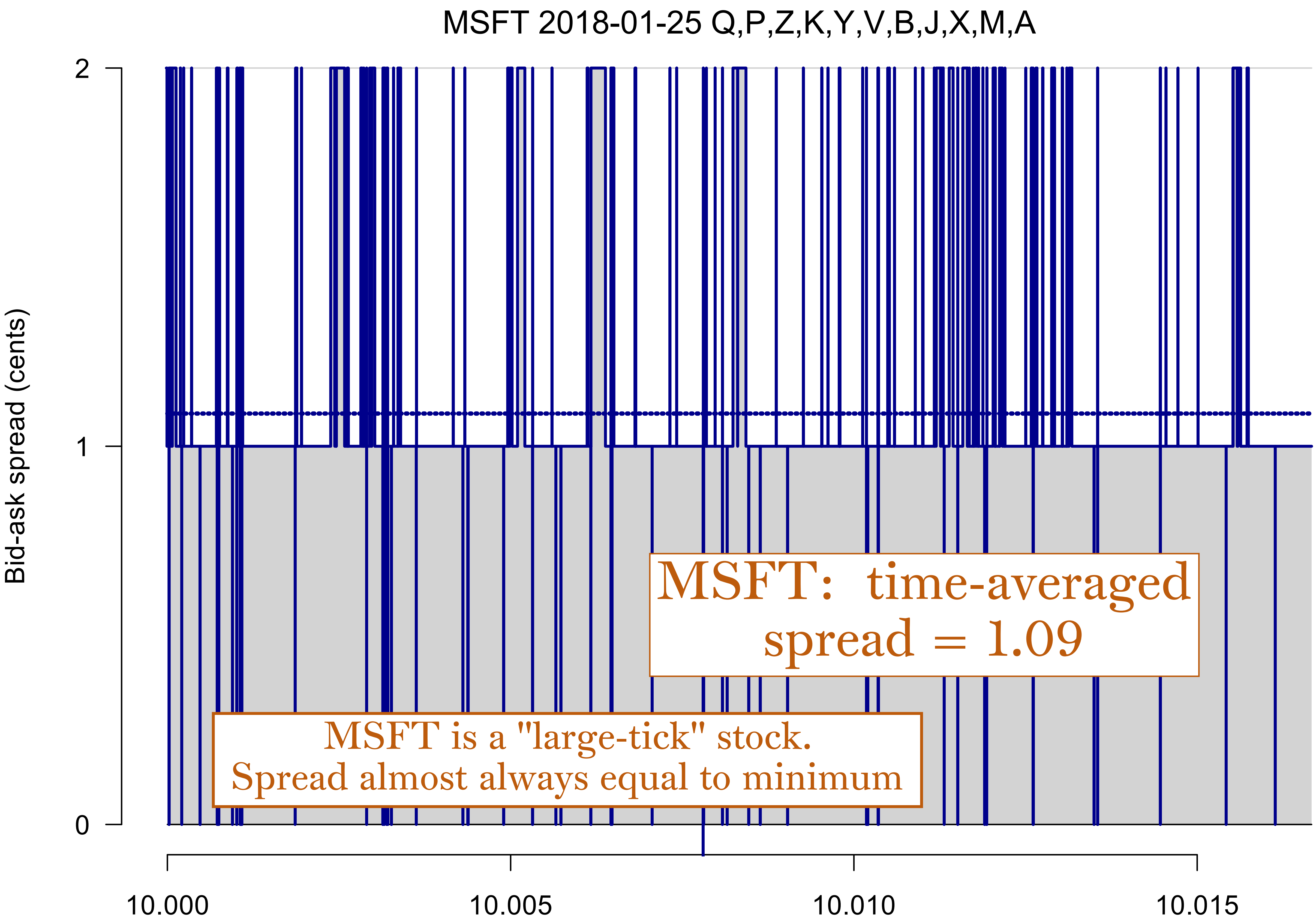
- What is "a trade"?

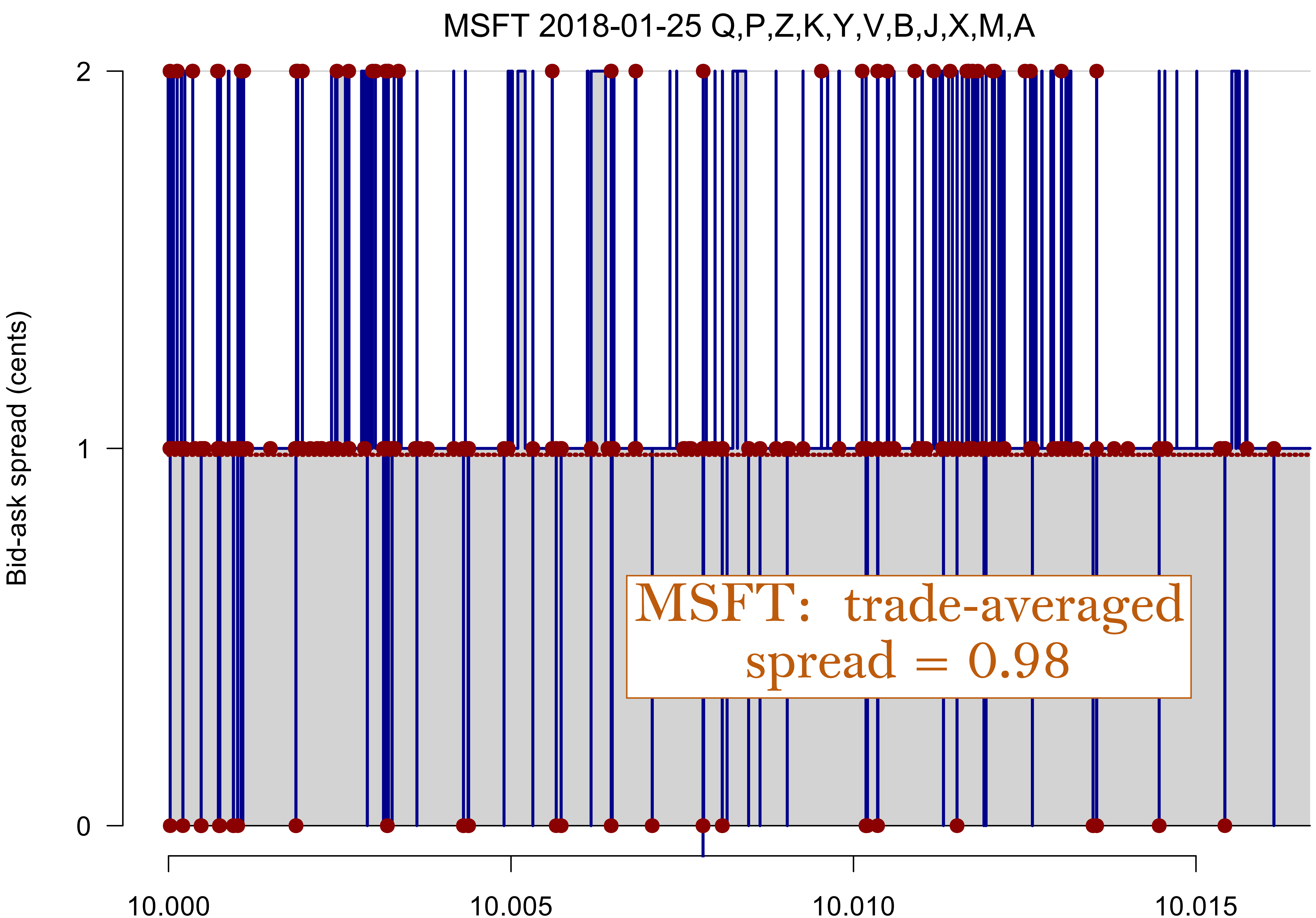
  Many (60%) separate prints may be part of same event
- Should you weight by trade size?

  Does one trade of 1000 shares = 10 x 100 shares

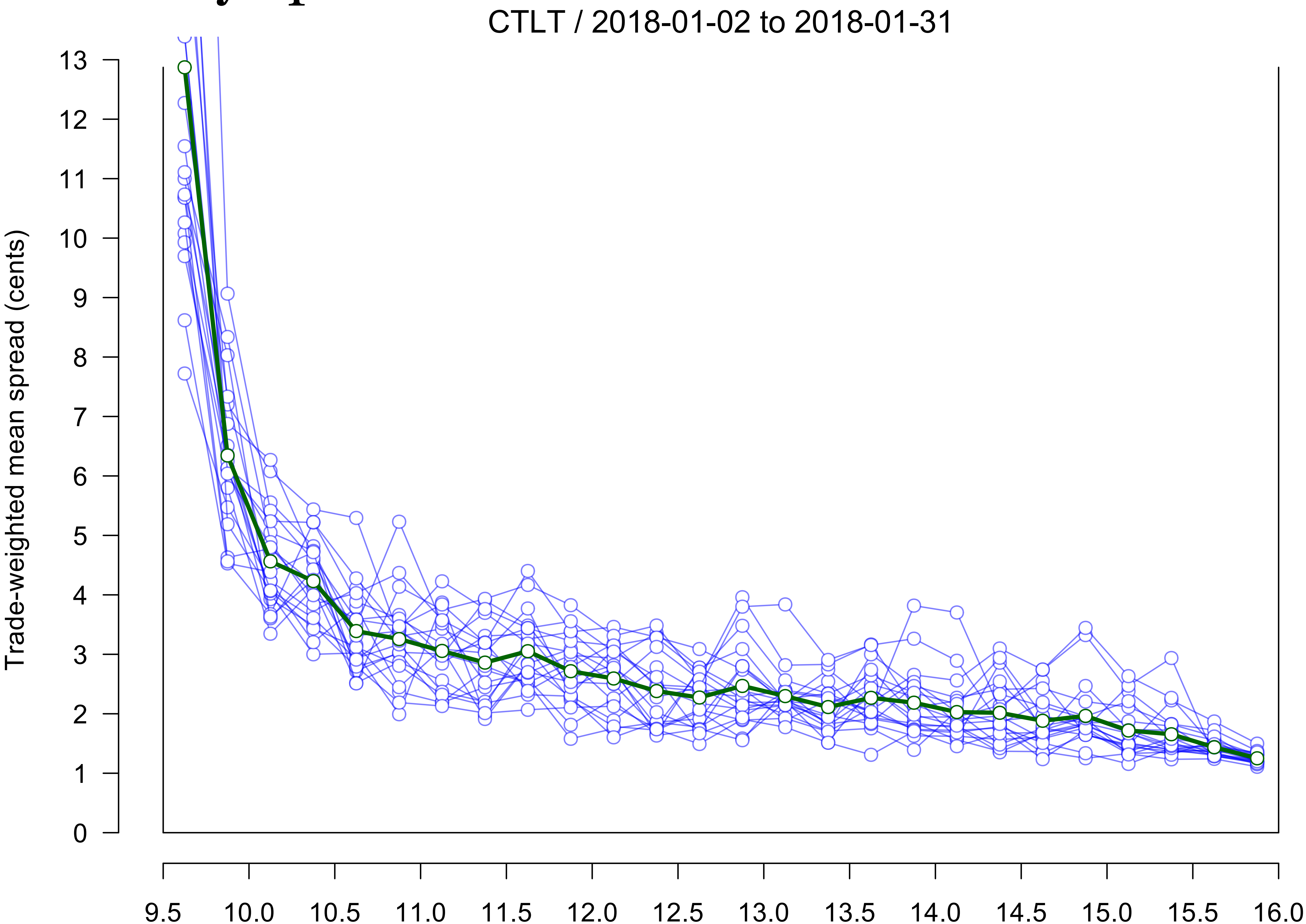
  What is the right measure of "activity"?

Not all spreads are greater than 1 tick




MSFT 2018-01-25 Q,P,Z,K,Y,V,B,J,X,M,A
Bid-ask spread (cents)
2
1
0
10.000
10.005
10.010
10.015
MSFT: trade-averaged spread = 0.98

Intraday spread curve



Use of intraday spread curve

- Forecast of liquidity

  small spread = more liquid = lower trading costs
- Also should forecast size at inside quotes

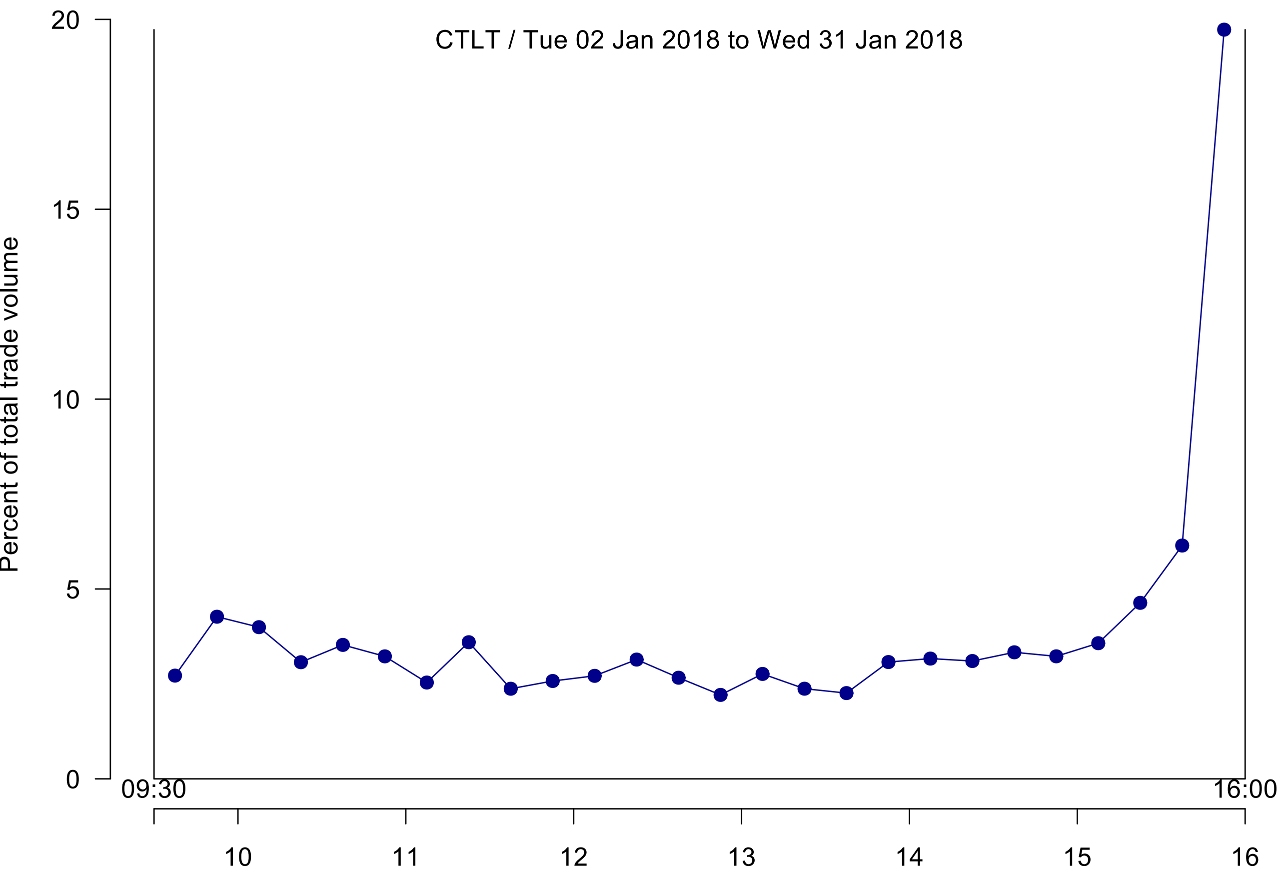
  especially for assets whose spread is always one tick

  complicated to evaluate for multi-tick assets

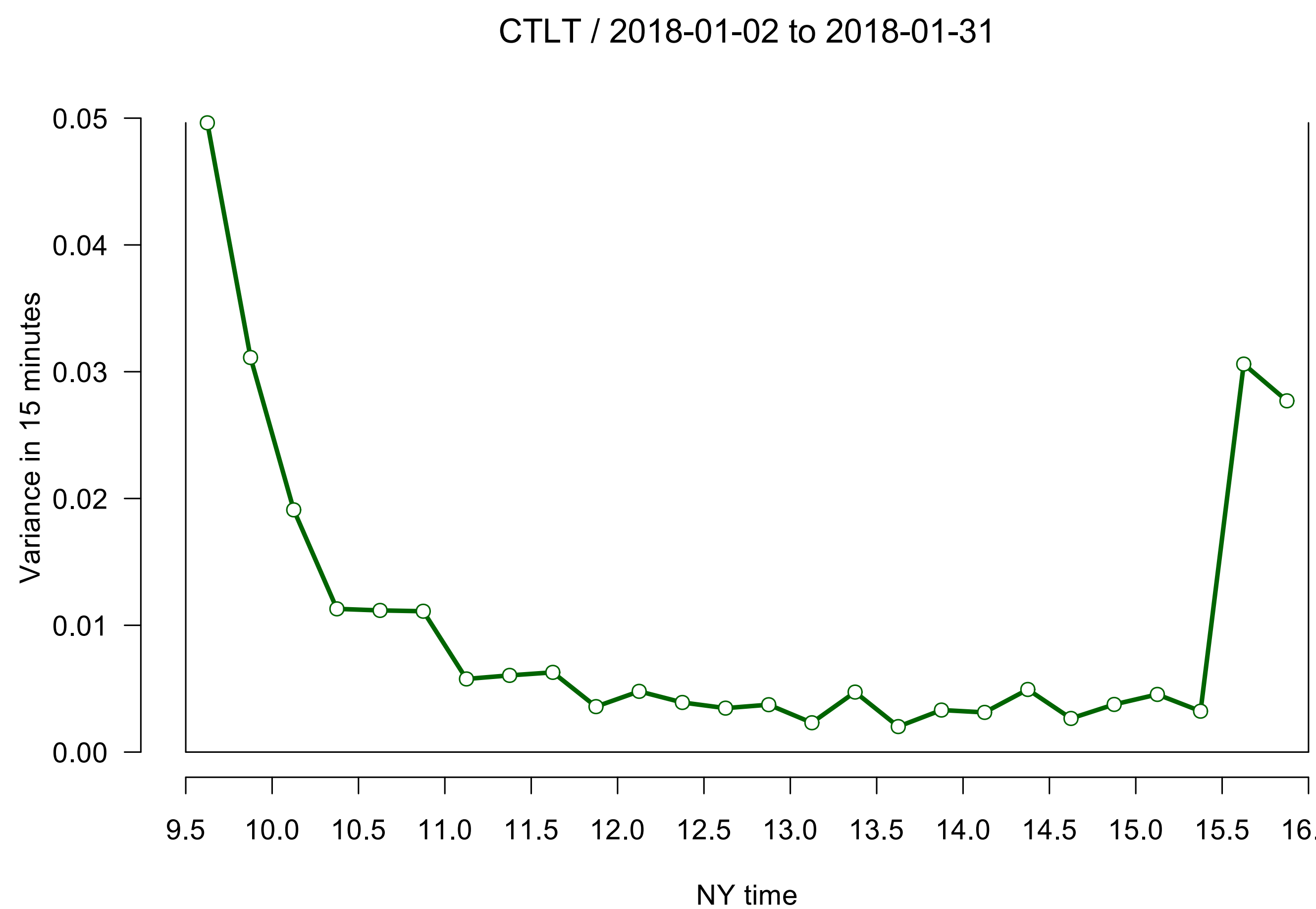
Volatility

- Why is it significant?
- How do you think about historical vs forecast?
- How do you model daily volatility?
- What changes when you go to intraday time scales?
- Construction of intraday volatility profiles.

Intraday volume profile



Intraday variance profile



How do we compute this profile of variance (volatility) and what does it mean?

FINANCIAL TIMES

The Big Read Markets volatility + Add to myFT

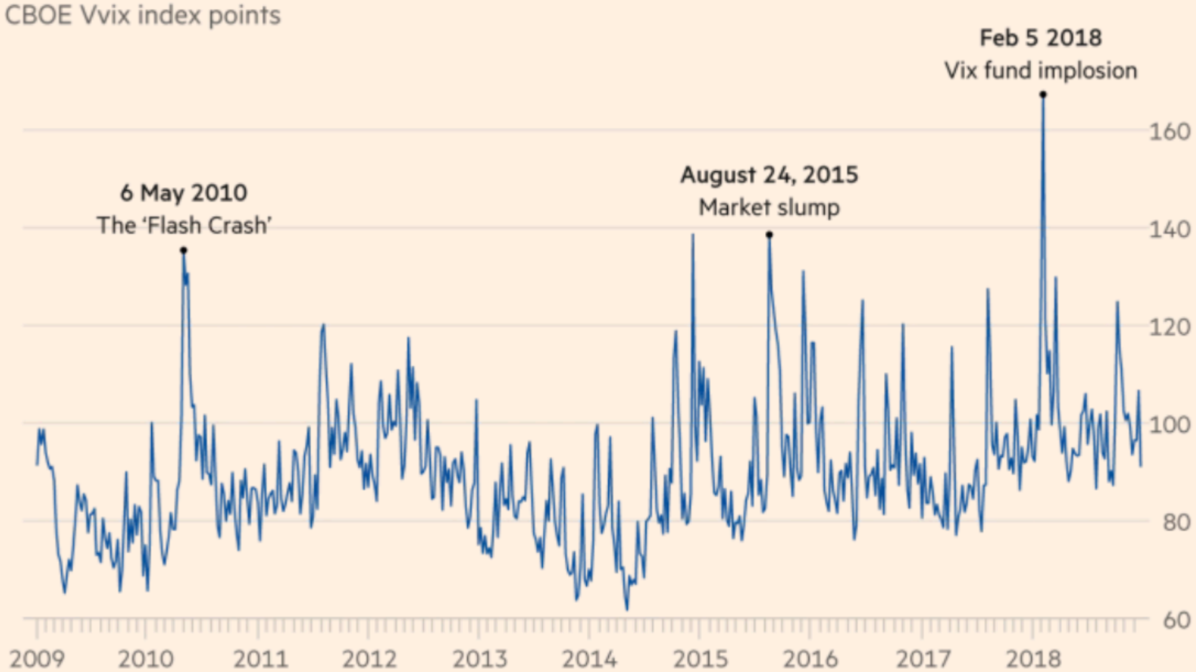
Volatility: how 'algos' changed the rhythm of the market

Critics say high-frequency trading makes markets too fickle amid rising anxiety over the global economy

Robin Wigglesworth in New York JANUARY 9, 2019

Even the volatility of volatility has been spiking lately

CBOE Vvix index points



Source: Bloomberg

Many investors and analysts blamed algorithmic strategies that automatically adjust their market exposure according to volatility for aggravating the 2015 crash. Targeting a specific level of volatility is common among strategies known as "risk parity" — trend-following hedge funds and "managed volatility" products sold by insurance companies. Estimates vary, but there is probably more than $1tn invested in a variety of such funds.

Risk parity, a strategy first pioneered by Ray Dalio's Bridgewater Associates in the 1990s, often shoulders much of the opprobrium. The theory is that a broad, diversified portfolio of stocks, bonds and other assets balanced by the mathematical risk — in practice, volatility — of each asset class should over time enjoy better returns than traditional portfolios. Bonds are less volatile than equities, so that often means "leveraging" these investments to bring the risk-adjusted allocation up to that of stocks. As volatility goes up, risk parity funds in theory rein in their exposure.

FINANCIAL TIMES

Markets volatility + Add to myFT

Volatility specialists face year of rewards and reckoning

Some funds say markets are not prepared for unwinding of the 'low vol' regime

Robin Wigglesworth JANUARY 3, 2019

Volatility returned with a bang last year, wrongfooting many investors but leaving traders that specialise in surfing the undulations of financial markets eagerly awaiting a new era of turbulence in 2019.

January started with a market melt-up that unravelled in dramatic fashion, when " short-volatility" bets that everything would remain calm came unstuck in spectacular fashion. Two popular short-volatility funds were shredded, exacerbating the stock market turmoil and shaking investors from their torpor.

"After years of relatively smooth sailing in the financial markets, heightened volatility is rocking the boat once again," Capital Group, the $1.9tn mutual fund manager, noted to clients. "There are many factors, but the three Ts are among the most impactful: tightening monetary policy, trade tensions and too much debt."

Benjamin Bowler, head of global equity derivatives research at Bank of America, argues that the "low vol bubble" that expanded in the post-crisis era is now deflating, and that many investors are still unprepared for a new more turbulent environment that is now beckoning.

"Looking at the world through the lens of volatility, we see markets that are unsustainably out of sync, fragile, and overall underpricing the risk of regime change," Mr Bowler notes. "History strongly suggests that the market we see today won't exist by the end of next year. While some markets have begun to 'wake up' to an inherently riskier world — less supported by central banks — many have not."

THE WALL STREET JOURNAL.
U.S. Edition ▾ | February 12, 2019 | Print Edition | Video

MARKETS

The Benchmark Set to Replace Libor Suffers Volatility Spike

If SOFR proves unusually volatile or hard to predict, it would diminish the rate's appeal to companies considering tying borrowing costs to it

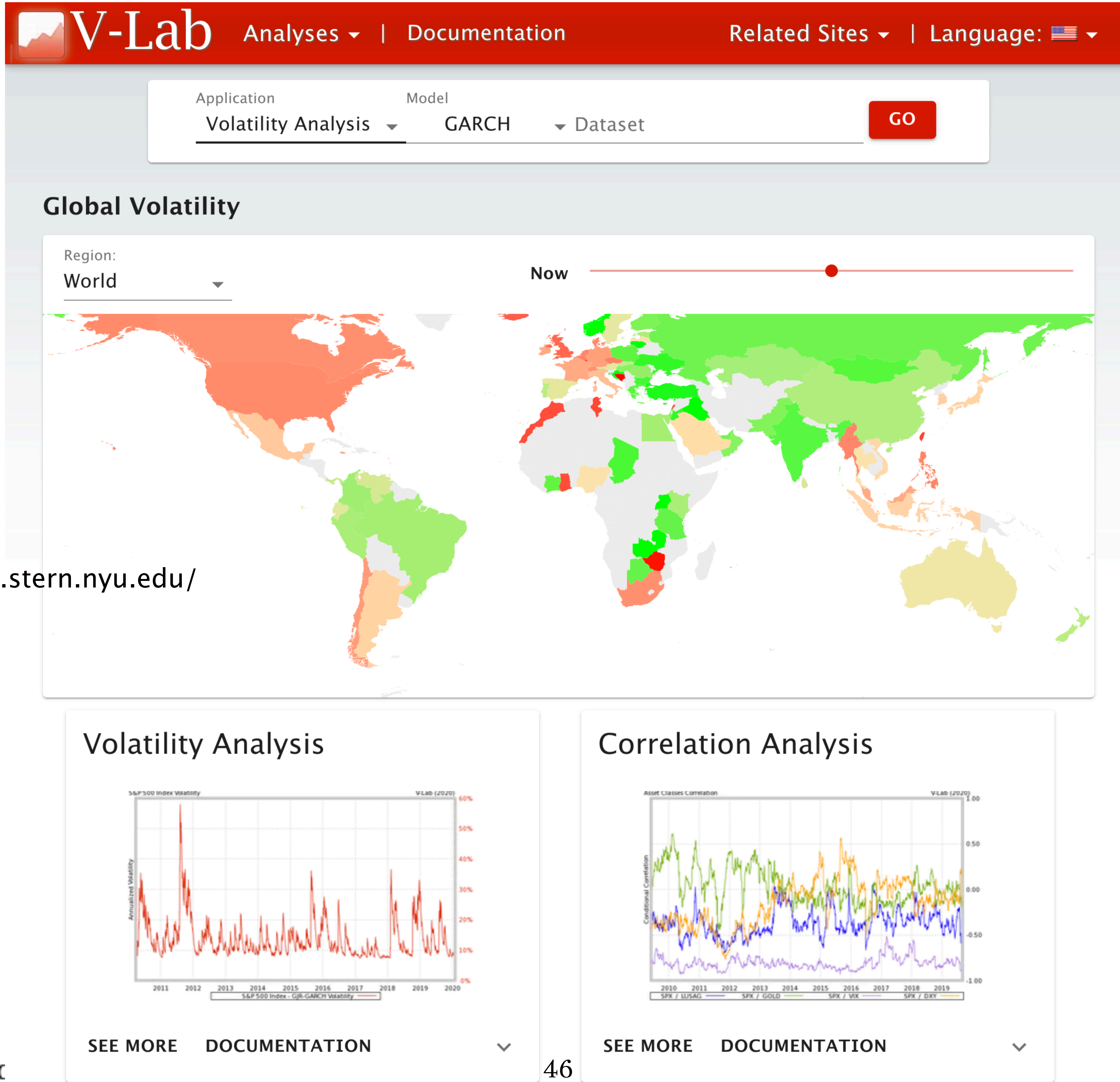
By *Daniel Kruger* and *Telis Demos* 9 COMMENTS

Updated Feb. 11, 2019 12:06 p.m. ET

Recent volatility in the market for overnight cash loans is raising concerns about a new benchmark that could set interest rates for trillions of dollars in mortgages and corporate debt.

Volatility =unpredictable price moves

Overnight repo rates can often be volatile, but the underlying reasons for and timing of many of those swings "are well-known and predictable," according to an explanation about SOFR on the website maintained by ARRC. When averaged over time, Treasury repo rates are less volatile than U.S. dollar Libor, the statement said.

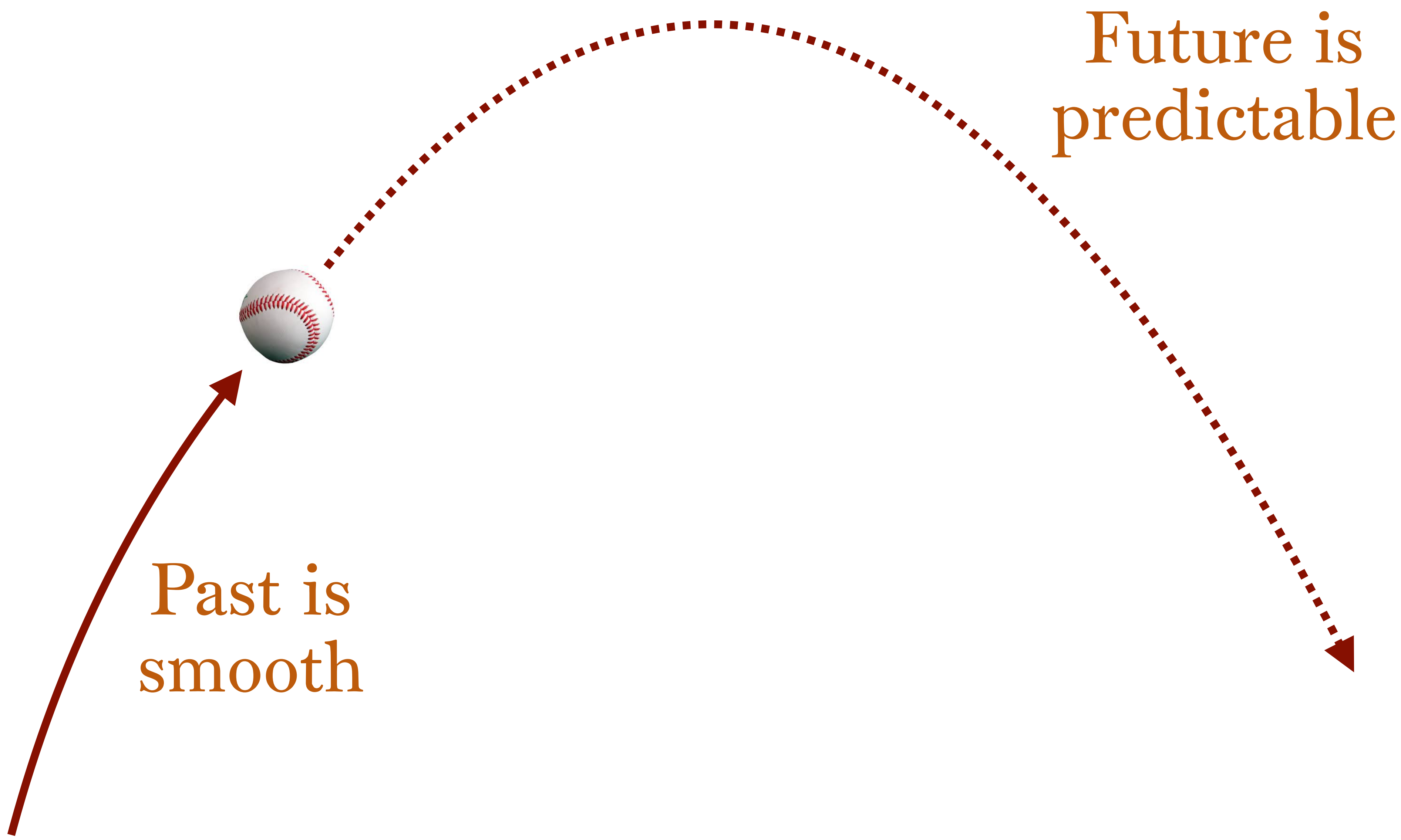


https://vlab.stern.nyu.edu/

Foundation of volatility

- History is "wiggly"
- Future is uncertain

Contrast to mechanical systems



A

BRIEF ACCOUNT

OF

MICROSCOPICAL OBSERVATIONS

Made in the Months of June, July, and August, 1827,

ON THE PARTICLES CONTAINED IN THE POLLEN OF PLANTS;

AND

ON THE GENERAL EXISTENCE OF ACTIVE MOLECULES

IN ORGANIC AND INORGANIC BODIES.

By ROBERT BROWN,

F.R.S., Hon. M.R.S.E. & R.I. Acad., V.P.L.S.

MEMBER OF THE ROYAL ACADEMY OF SCIENCES OF SWEDEN, OF THE ROYAL SOCIETY OF DENMARK, AND OF THE IMPERIAL ACADEMY NATURÆ CURIOSORUM; CORRESPONDING MEMBER OF THE ROYAL INSTITUTES OF FRANCE AND OF THE NETHERLANDS, OF THE IMPERIAL ACADEMY OF SCIENCES AT ST. PETERSBURG, AND OF THE ROYAL ACADEMIES OF PRUSSIA AND BAVARIA, &c.

While examining the form of these particles immersed in water, I observed many of them very evidently in motion; their motion consisting not only of a change of place in the fluid, manifested by alterations in their relative positions, but also not unfrequently of a change of form in the particle itself; a contraction or curvature taking place repeatedly about the middle of one side, accompanied by a corresponding swelling or convexity on the opposite side of the particle. In a few instances the particle was seen to turn on its longer axis. These motions were such as to satisfy me, after frequently repeated observation, that they arose neither from currents in the fluid, nor from its gradual evaporation, but belonged to the particle itself.

INVESTIGATIONS ON THE THEORY OF THE BROWNIAN MOVEMENT

BY

ALBERT EINSTEIN, Ph.D.

EDITED WITH NOTES BY

R. FÜRTH

TRANSLATED BY

A. D. COWPER

This Dover edition, first published in 1956, is an unabridged and unaltered republication of the translation originally published in 1926. It is published through special arrangement with Methuen and Co., Ltd., and the estate of Albert Einstein.

Evidently it must be assumed that each single particle executes a movement which is independent of the movement of all other particles; the movements of one and the same particle after different intervals of time must be considered as mutually independent processes, so long as we think of these intervals of time as being chosen not too small.

L. BACHELIER
Théorie de la spéculation
Annales scientifiques de l'É.N.S. 3e série, tome 17 (1900), p. 21-86.
<http://www.numdam.org/item?id=ASENS_1900_3_17__21_0>

THÉORIE DE LA SPÉCULATION,

PAR M. L. BACHELIER.

INTRODUCTION.

Les influences qui déterminent les mouvements de la Bourse sont innombrables, des événements passés, actuels ou même escomptables, ne présentant souvent aucun rapport apparent avec ses variations, se répercutent sur son cours.

A côté des causes en quelque sorte naturelles des variations, interviennent aussi des causes factices : la Bourse agit sur elle-même et le mouvement actuel est fonction, non seulement des mouvements antérieurs, mais aussi de la position de place.

La détermination de ces mouvements se subordonne à un nombre infini de facteurs : il est dès lors impossible d'en espérer la prévision mathématique. Les opinions contradictoires relatives à ces variations se partagent si bien qu'au même instant les acheteurs croient à la hausse et les vendeurs à la baisse.

Le Calcul des probabilités ne pourra sans doute jamais s'appliquer aux mouvements de la cote et la dynamique de la Bourse ne sera jamais une science exacte.

Mais il est possible d'étudier mathématiquement l'état statique du marché à un instant donné, c'est-à-dire d'établir la loi de probabilité des variations de cours qu'admet à cet instant le marché. Si le marché, en effet, ne prévoit pas les mouvements, il les considère comme étant

truisent donc deux à deux et il ne reste que la dérivée du dernier par rapport à p

$$\frac{m!}{mp!\,mq!}p^{mp}q^{mq}mpq.$$

La valeur moyenne de l'écart h serait égale au double de cette quantité.

Lorsque le nombre m est suffisamment grand, on peut simplifier les expressions précédentes en faisant usage de la formule asymptotique de Stirling

$$n! = e^{-n}n^n\sqrt{2\pi n}.$$

On obtient ainsi pour l'espérance mathématique la valeur

$$\frac{\sqrt{mpq}}{\sqrt{2\pi}}.$$

La probabilité pour que l'écart h soit compris entre h et $h+dh$ aura pour expression

$$\boxed{\frac{dh}{\sqrt{2\pi mpq}}e^{-\frac{h^2}{2mpq}}.}$$

Nous pouvons appliquer la théorie qui précède à notre étude. On peut supposer le temps divisé en intervalles très petits Δt; de sorte que $t = m\Delta t$; pendant le temps Δt le cours variera probablement très peu.

Formons la somme des produits des écarts qui peuvent exister à l'époque Δt par les probabilités correspondantes; c'est-à-dire $\int_0^\infty px\,dx$, p étant la probabilité de l'écart x.

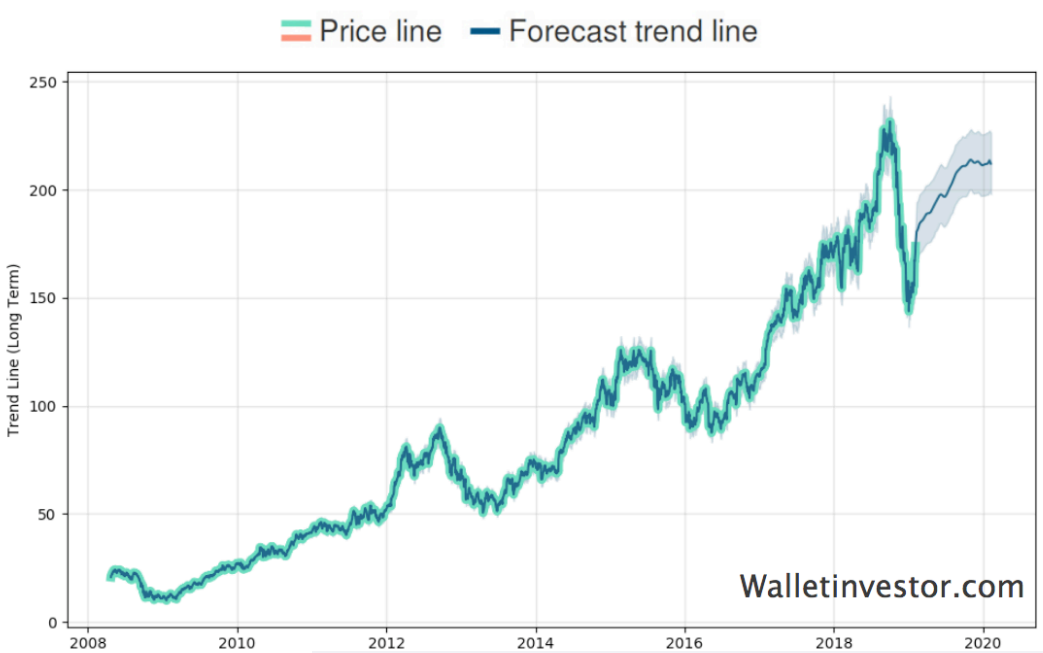
Cette intégrale doit être finie, car, par suite de la petitesse supposée de Δt, les écarts considérables ont une probabilité évanouissante. Cette intégrale exprime du reste une espérance mathématique, qui ne peut être finie si elle correspond à un intervalle de temps très petit.

Désignons par Δx le double de la valeur de l'intégrale ci-dessus; Δx sera la moyenne des écarts ou l'écart moyen pendant le temps Δt.

Si le nombre m des épreuves est très grand et si la probabilité reste la même à chaque épreuve, nous pourrons supposer que le cours varie

Daily closing price for AAPL (unadjusted for dividends)

Future prices are uncertain



Forecast for AAPL - Based on Mon Feb 11 2019

Share price forecast

The 34 analysts offering 12 month price targets for Apple Inc. have a median target of 177.99, with a high estimate of 245.00 and a low estimate of 140.00. The median estimate represents a 5.05% increase from the last price of 169.43.

High	44.6%	245.00
Med	5.1%	177.99
Low	-17.4%	140.00

Warning: Correction likely over. However, risks factor like high corporate debt and slower global growth remain to be a concern. High volatility likely to stay around.

5 Trading Days Later - Tue Feb 19 2019 (more indicators)

Status: Might Go Up

Number of data points: 55 (31 clusters)

Increase probability: 67.88% (± 14.28%)

Return: 2.12 ± 6.40 % ($173.03 ± 10.84)

Indicators used: RSI around 59.78, Middle Keltner Slope increasing, Upper-to-Lower-Keltner around 1.10

10 Trading Days Later - Tue Feb 26 2019 (more indicators)

Status: Might Go Up

Number of data points: 304 (112 clusters)

Increase probability: 70.67% (± 5.92%)

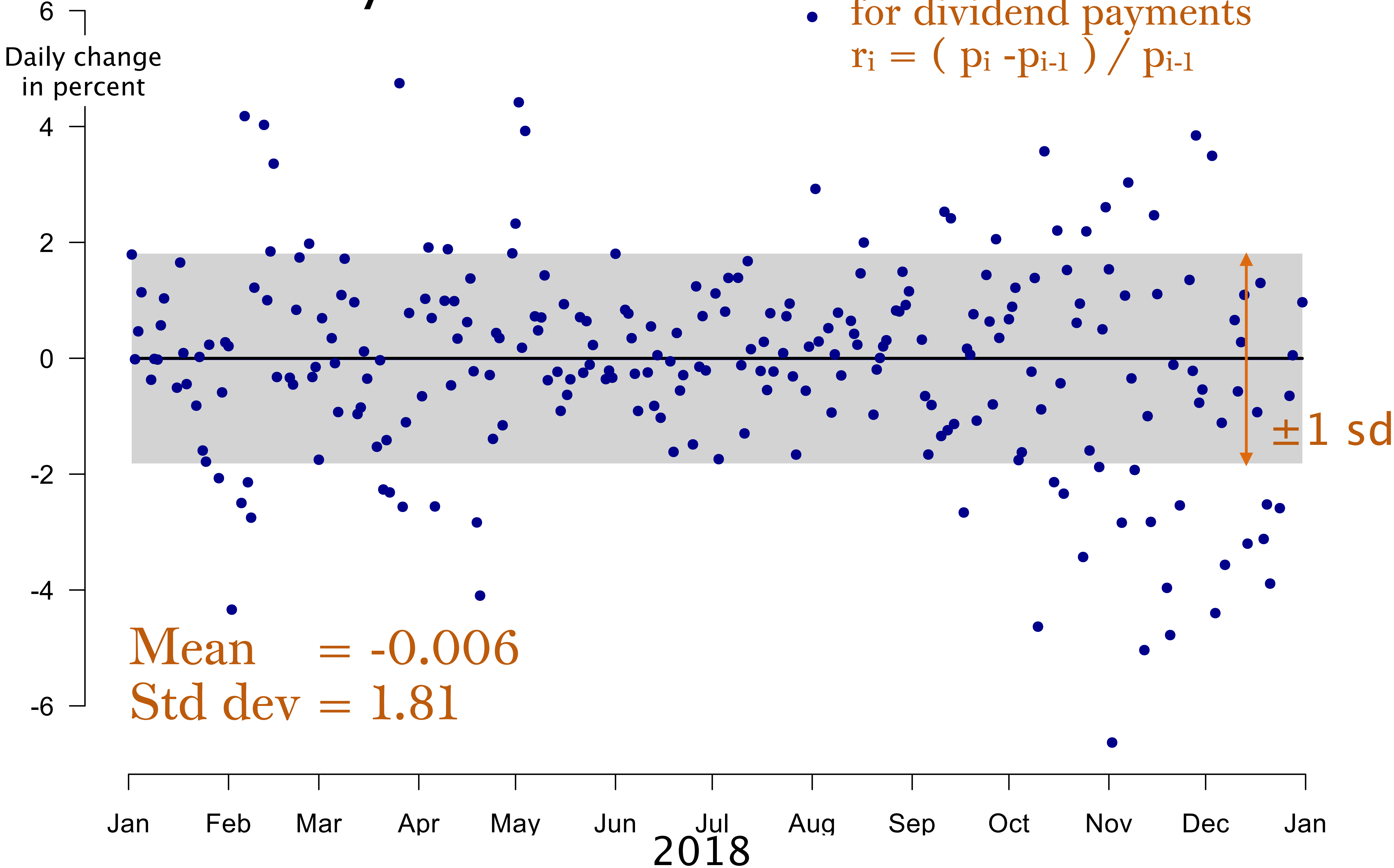
Return: 2.60 ± 7.56 % ($173.84 ± 12.81)

Indicators used: Price-to-Lower-Keltner around 1.09, Middle Keltner Slope increasing, SMA200 Slope increasing

Stochastic model

- Make a *probabilistic* model for future price changes based on *statistics* of past price changes
- *Expected value* of future change is essentially zero (efficient market hypothesis) except for small info
- *Variance* of future changes (as probability) can be well predicted by variance of past changes (sample)

AAPL 2018: 250 daily returns



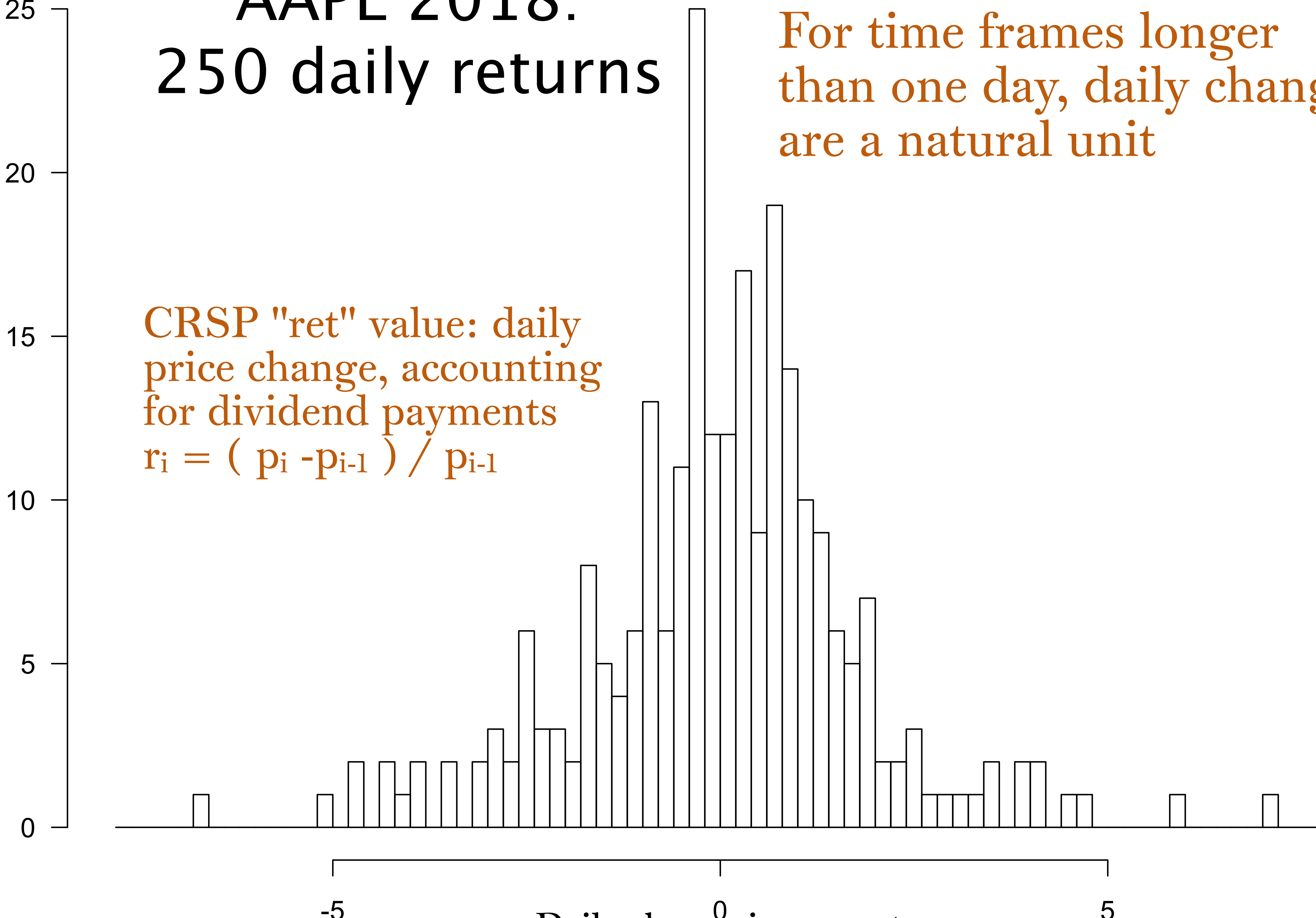
CRSP "ret" value: daily price change, accounting for dividend payments
$r_i = (p_i - p_{i-1}) / p_{i-1}$

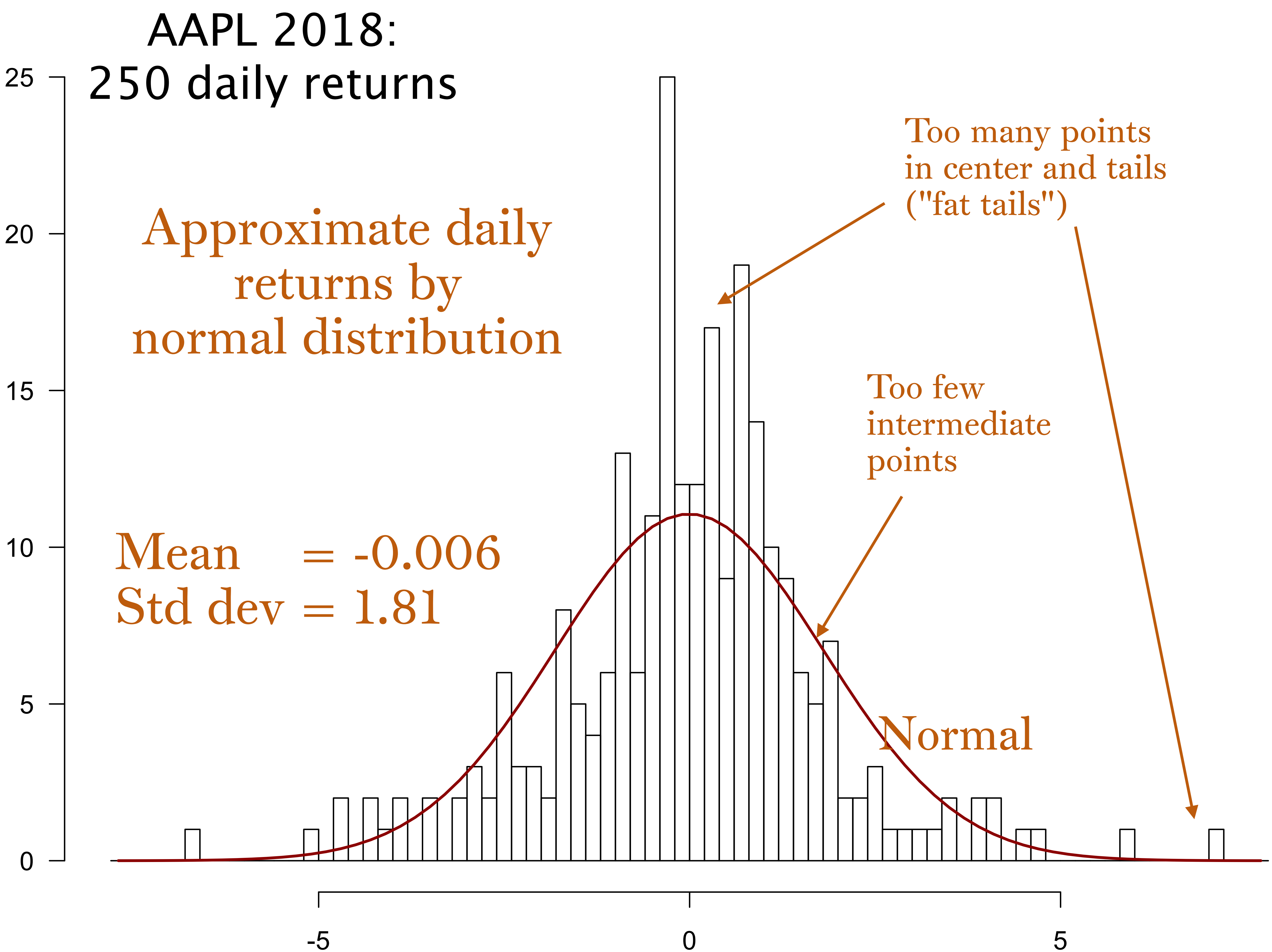
Mean = -0.006
Std dev = 1.81

AAPL 2018: 250 daily returns

For time frames longer than one day, daily changes are a natural unit

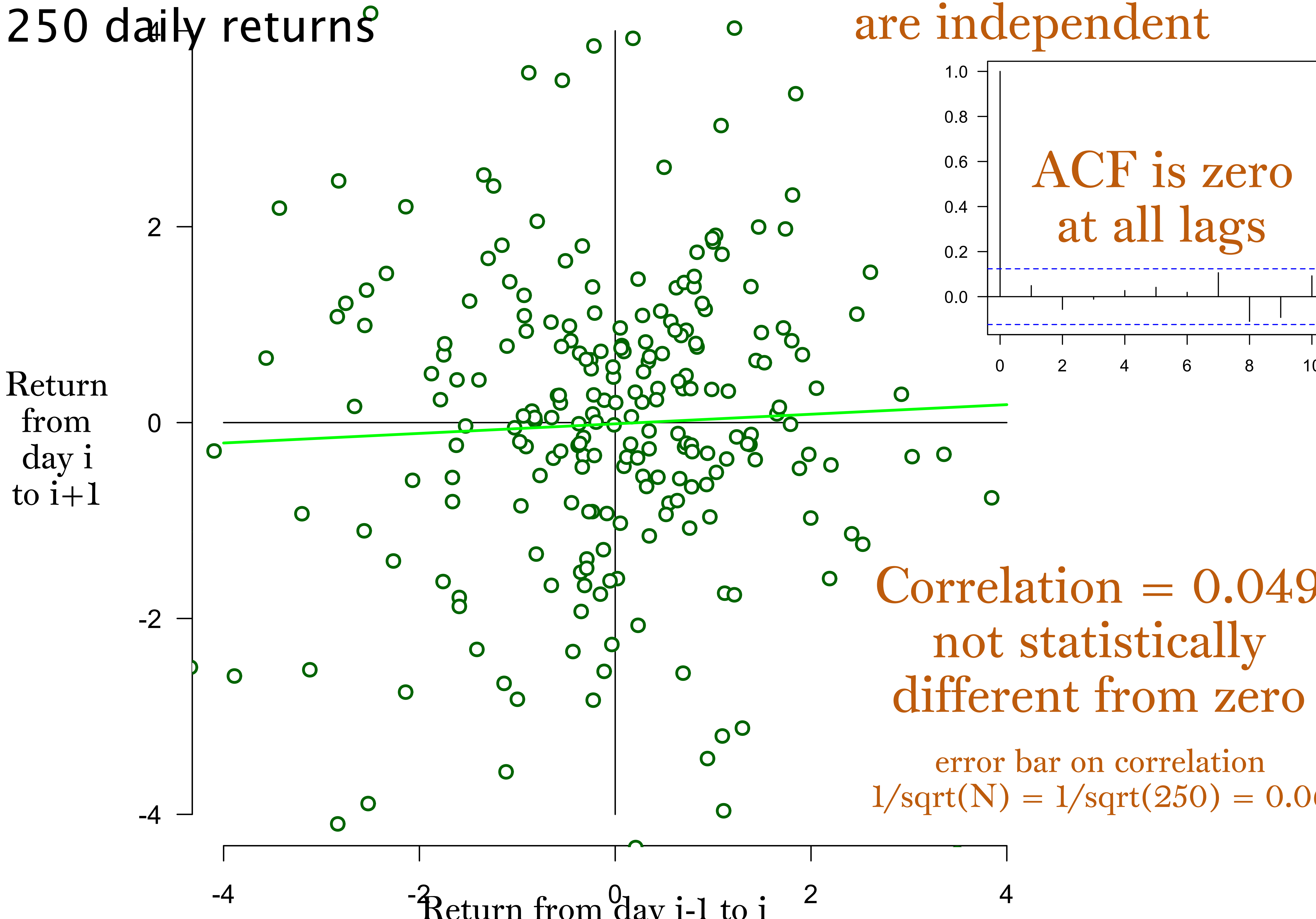
CRSP "ret" value: daily price change, accounting for dividend payments
$r_i = (p_i - p_{i-1}) / p_{i-1}$




AAPL 2018:
250 daily returns
Approximate daily returns by normal distribution
Mean = -0.006
Std dev = 1.81
Too many points in center and tails ("fat tails")
Too few intermediate points
Normal
25
20
15
10
5
0
-5
0
5

AAPL 2018:
250 daily returns

Consecutive daily returns are independent



ACF is zero at all lags

Correlation = 0.049 not statistically different from zero

error bar on correlation
1/sqrt(N) = 1/sqrt(250) = 0.06

Model for daily returns

Historical data is more or less consistent with model

$$
\begin{aligned}
r_i &= (p_i - p_{i-1})/p_{i-1} \\
&\approx \log p_i - \log p_{i-1} \\
&\sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

(price changes are small)

(mean of daily price changes is much smaller than standard deviation)

or

$$
\begin{aligned}
\log p_i &= \log p_{i-1} + \sigma \xi_i \\
\sigma &= \text{volatility}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(\xi_i) &= 0 \\
\mathrm{Var}(\xi_i) &= 1
\end{aligned}
$$

ξ_i, ξ_j independent for $i \neq j$

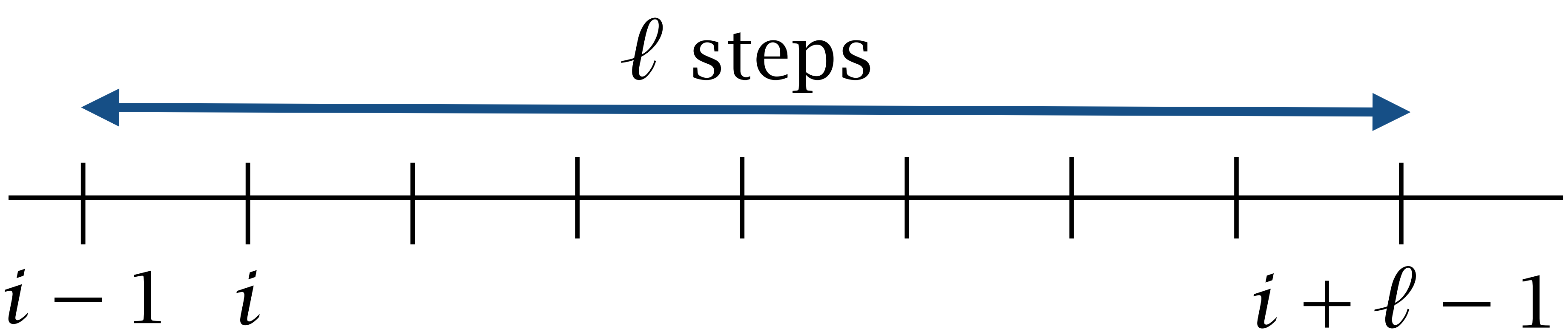
For example, $\xi_i \sim \mathcal{N}(0, 1)$

Volatility is mean squared daily fractional price change.

Daily model also tells you statistics on longer intervals

$$\log p_i \;=\; \log p_{i-1} \;+\; \sigma\,\xi_i$$

$$\log p_{i+\ell-1} \;=\; \log p_{i-1} \;+\; \sigma \sum_{j=i}^{i+\ell-1} \xi_j$$



$$\sigma \sum_{j=i}^{i+\ell-1} \xi_j \;\sim\; \mathcal{N}\left(0,\, \ell\sigma^2\right)$$

because of
serial independence,
and for large lag

Volatility squared = variance per day

Units of σ are day$^{-1/2}$

Assumptions of model

- Log price does a Brownian motion
- Drift is negligible
- Volatility **σ** is constant:

  no random variation

  no cyclical structure

Then forecast volatility is historical mean

All these assumptions can be criticized

Variations in volatility

- Volatility is an asset class like any other
- Volatility may have predictable patterns

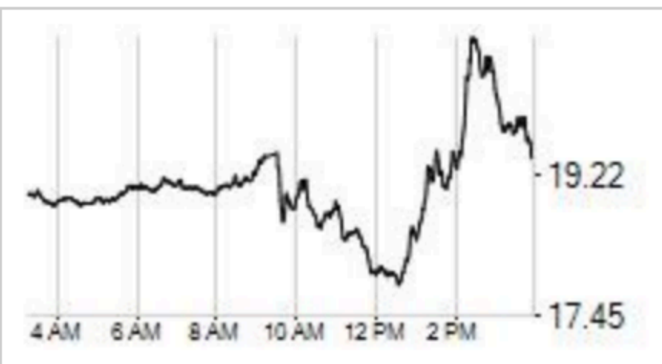
Cboe

U.S. Options | U.S. Equities | European Equities | Futures | LiveVol | Indexes | FX

Quotes & Data News Products Education Strategies Tools & Resources About Cboe Institutional
Advisors Technology Integration

Cboe Volatility Index® (VIX®)

VIX



SPX	2732.22	1.02
VIX	19.46	0.33
08VX/G8	19.20	0.70
09VX/G8	18.90	0.65
10VX/H8	18.25	0.43

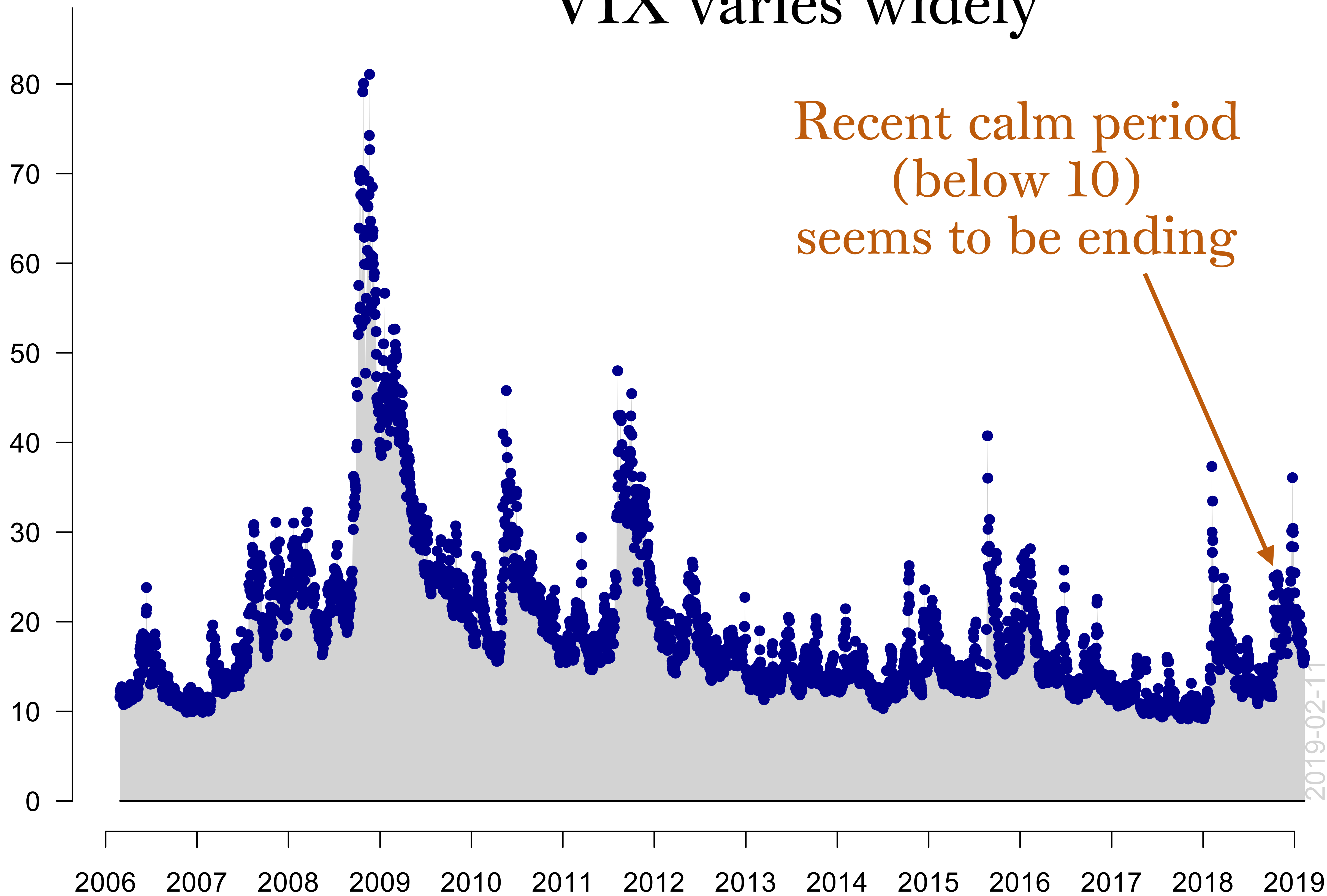
Delayed Quotes

Prices for 3 VIX **Futures** are above

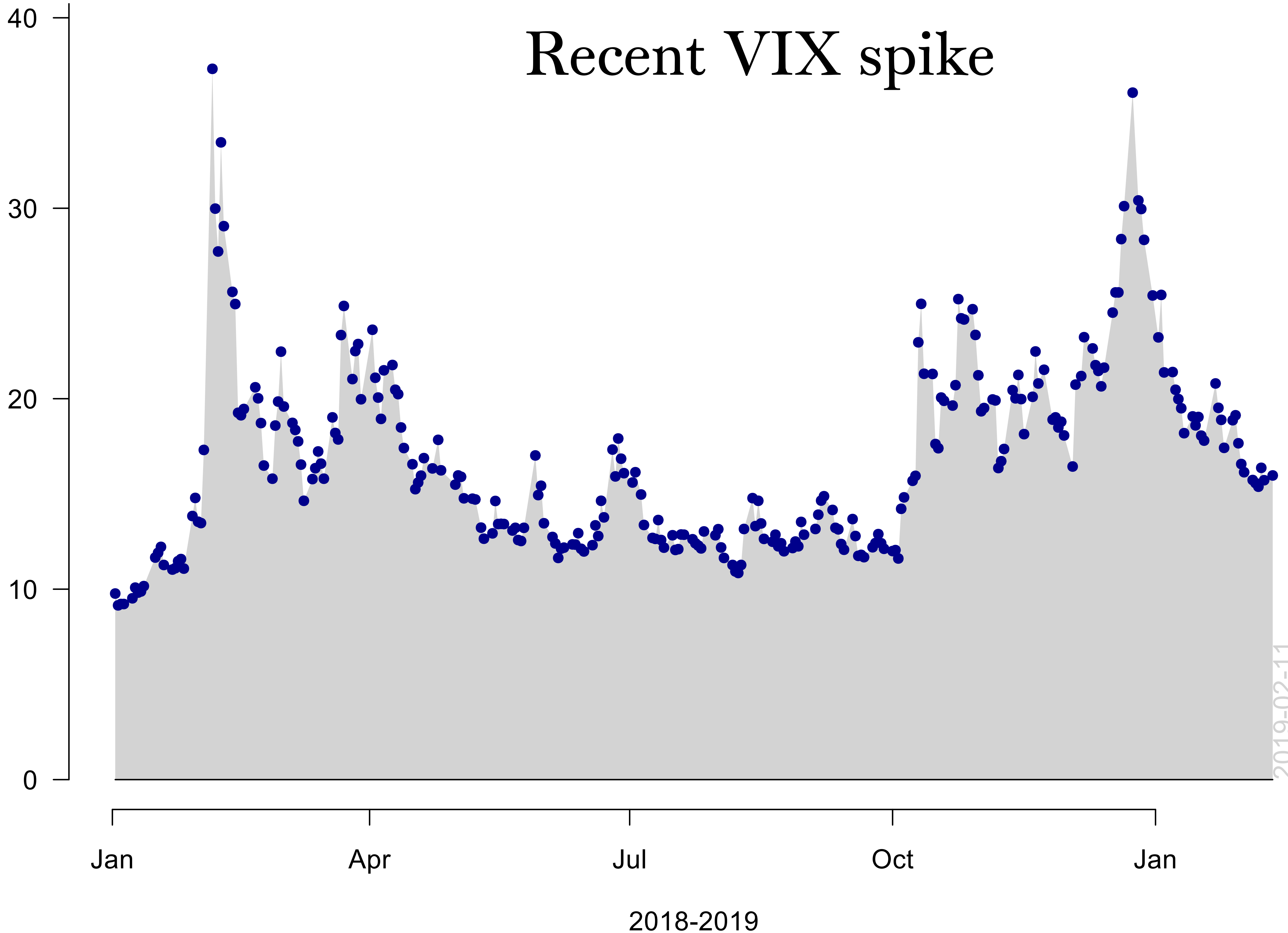
VIX is market belief about future volatility of equity market ("implied" volatility from options prices)

The Cboe Volatility Index® (VIX® Index®) is a key measure of market expectations of near-term volatility conveyed by S&P 500 stock index option prices. Since its introduction in 1993, the VIX Index has been considered by many to be the world's premier barometer of investor sentiment and market volatility. Several investors expressed interest in trading instruments related to the market's expectation of future volatility, and so VX futures were introduced in 2004, and VIX options were introduced in 2006.

VIX varies widely



Recent VIX spike



Volatility forecasting from time series

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003
Robert F. Engle III, Clive W.J. Granger

Share this:

KUNGL. VETENSKAPSAKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

English
Swedish

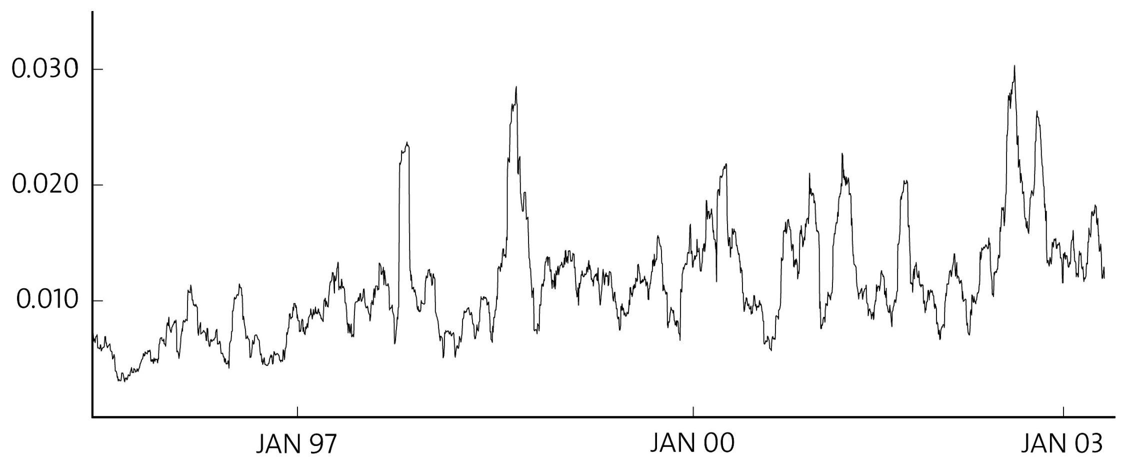


Figure 3: Standard deviation for percentage daily returns on an investment in the Standard & Poor 500 stock index, May 16, 1995–April 29, 2003, computed from data for the four preceding weeks.

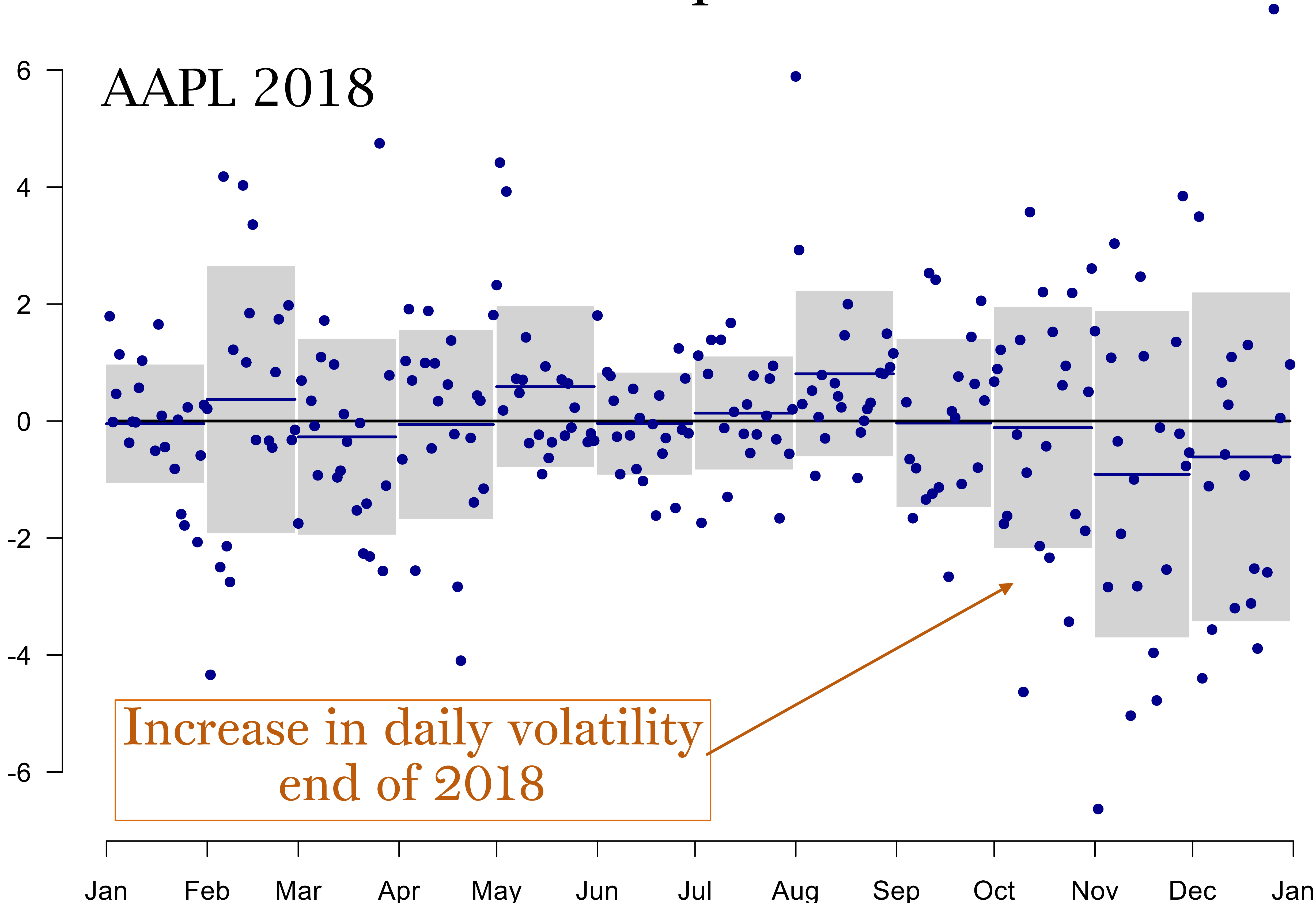
Engle's Contribution

Figure 3 shows backward-looking calculations of time-varying volatility. But investors and financial institutions need forward-looking evaluations – forecasts – of volatility during the next day, week and year. In an outstanding article in 1982, Robert Engle formulated a model which allows such evaluations.

Statistical models of asset returns can only explain a fraction of the variation from one day to the next. Most of the volatility is thus embedded in the random error term (e_t in the introductory equation) – or, in other words, in the model's forecasting error. In standard statistical models, the expected variance of the random error is assumed to be constant over time. Obviously, this is far from capturing the large variations in asset returns depicted in *Figure 3.*

Engle assumed instead that the variance of the random error in a certain statistical model, in a certain time period, systematically depends on previously realized random errors, so that large (small) errors tend to be followed by large (small) errors. In technical terms, the random variable displays *autoregressive conditional heteroskedasticity*. His approach has therefore become acronymized ARCH. In our example, the model now contains not only a forecasting equation for asset returns, but also a number of parameters showing how the variance of the random error in this equation depends on forecasting errors in earlier periods. Engle demonstrated how ARCH models could be estimated and introduced a practical test for the hypothesis that the conditional variance of the random error is constant.

Calendar profile?



Summary of daily volatility

- Calibrated from historical data,
  suggests stochastic model for future evolution
  - approximately normal daily returns
  - no serial dependence in daily returns
- Volatility often taken to be constant in model,
  though in reality certainly varies randomly
  (predictable annual profile is less convincing)
- Daily data is the natural finest resolution
  can build up longer intervals by aggregating daily

High-frequency volatility compared to daily

- There is no natural finest time scale

  data arrives in continuous time
- It is not obvious what "the price" is

  last trade price? bid? ask? midpoint?
- Serial independence is violated on short time scales

  require special statistical techniques to use fine data
- Intraday profiles cannot be neglected

Model for time evolution

(drop the logarithms -- not important on intraday time scales)

- Discrete time (daily data): Typically, constant volatility

difference eqn: $$p_i - p_{i-1} = \sigma \xi_i, \qquad \xi_i \sim \mathcal{N}(0,1)$$

summed: $$p_i = p_0 + \sigma \sum_{j=1}^{i} \xi_j$$

- Continuous time (intraday data):

Possibly time-varying volatility

stochastic differential eqn: $$dp(t) = \sigma(t)\, dB(t)$$

$B(t)$ = Brownian process

integrated: $$p(t) = p(0) + \int_0^t \sigma(s)\, dB(s)$$

Variance across time intervals

- Discrete time:

$$p_{i+\ell} - p_i \quad \text{has variance} \quad \sigma^2 \ell$$

$$\sigma^2 \quad \text{is variance per step}$$

- Continuous time:

$$p(T) - p(0) \quad \text{has variance} \quad \int_0^T \sigma(t)^2 \, dt$$

$$\sigma(t)^2 \quad \text{is variance per unit time}$$

Estimation from historical data

- Discrete time:

$$N\sigma^2 \approx \sum_{j=1}^{N} (p_j - p_{j-1})^2$$

- Continuous time:

"quadratic variation"

$$\int_0^T \sigma(t)^2\, ds \approx \sum_{j=1}^{N} \Big(p(t_j) - p(t_{j-1})\Big)^2$$

any points $\{t_j\}_{j=0}^{N}$ with

$$0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$$

What prices to use?

- Possible prices:
  - last trade price
  - last quote update whether bid or ask (really bad choice)
  - bid-ask midpoint
- Choose the one that gives smallest serial dependence on reasonable time scales
  - trade prices have strong serial covariance (Roll model)

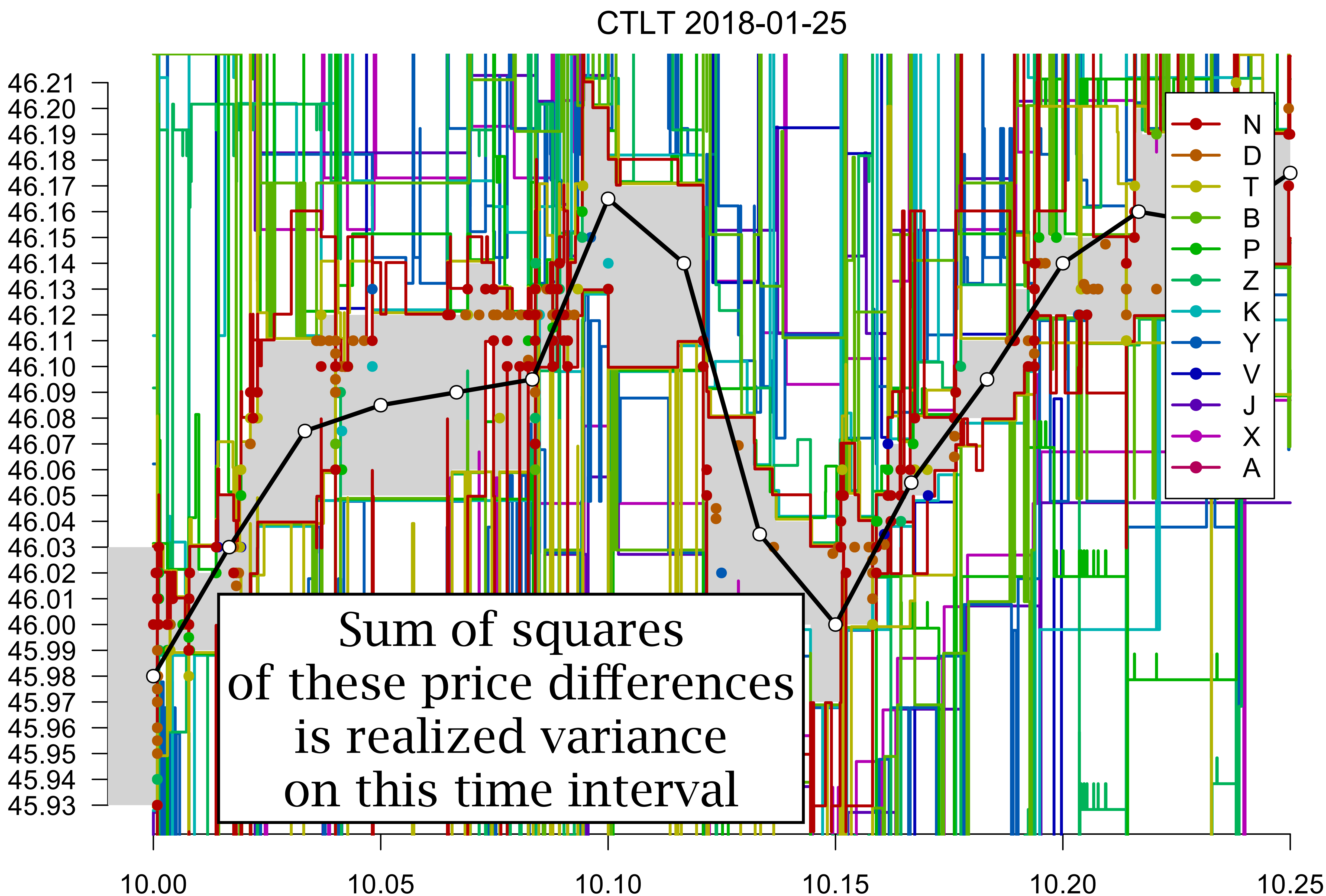
$\Longrightarrow$ Use bid-ask midpoint as "the price"

What times to use for sample?

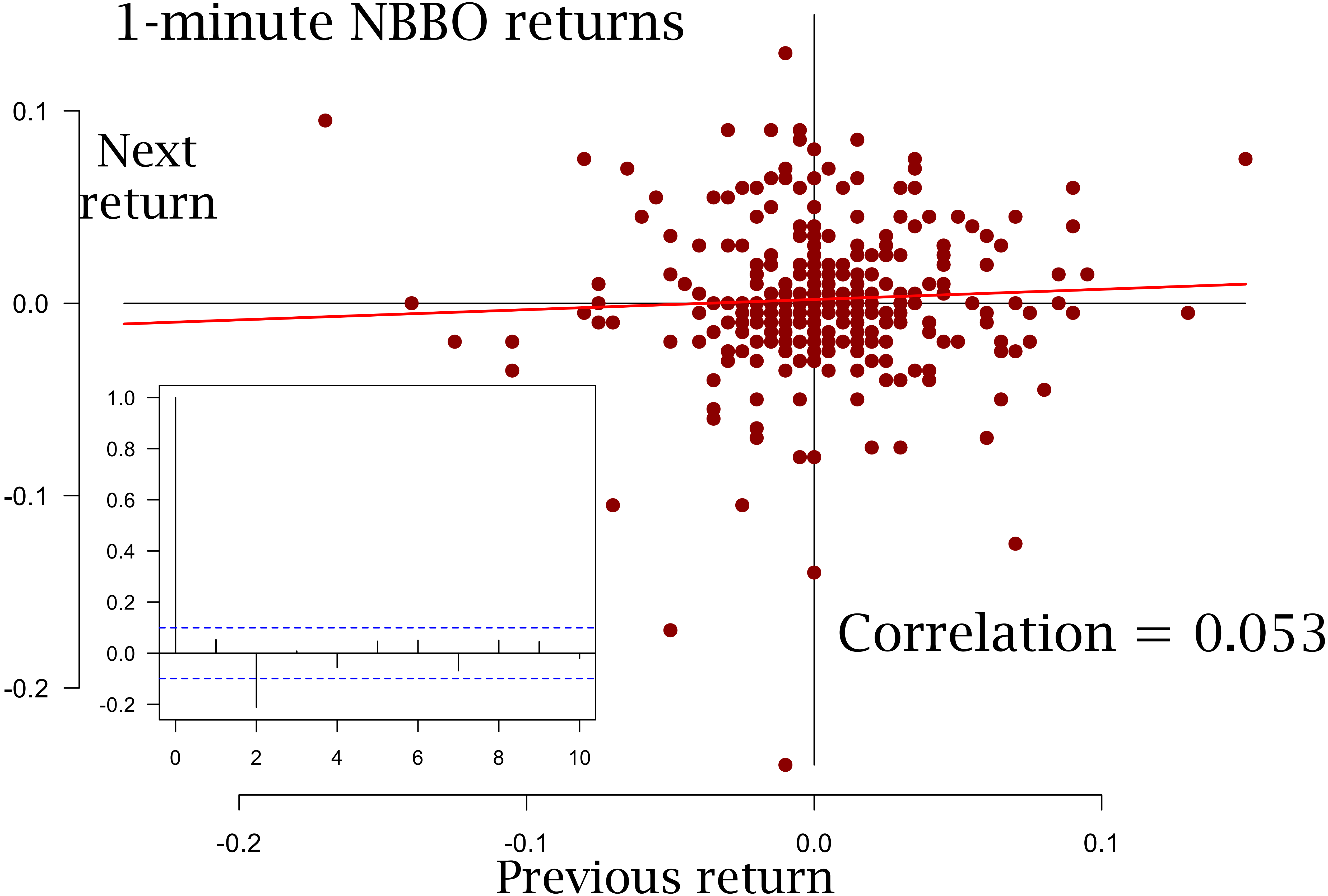
- Want to use fine times, to have lots of observations, and get good statistical precision
- Serial correlation is nontrivial at fine time scales

$\Longrightarrow$ 1-minute sampling is OK to start with

NBBO midpoint sampling at 1-minute intervals

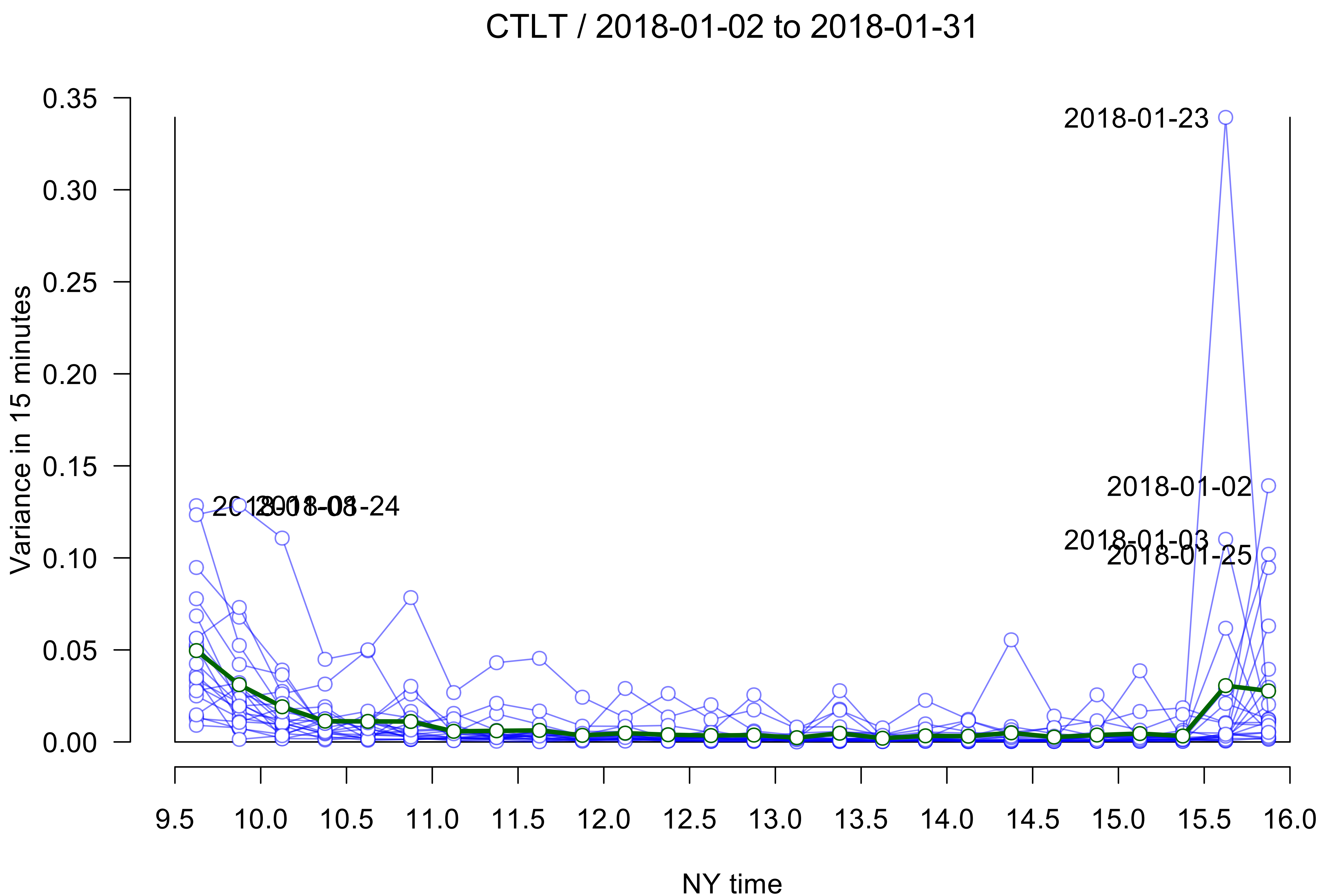


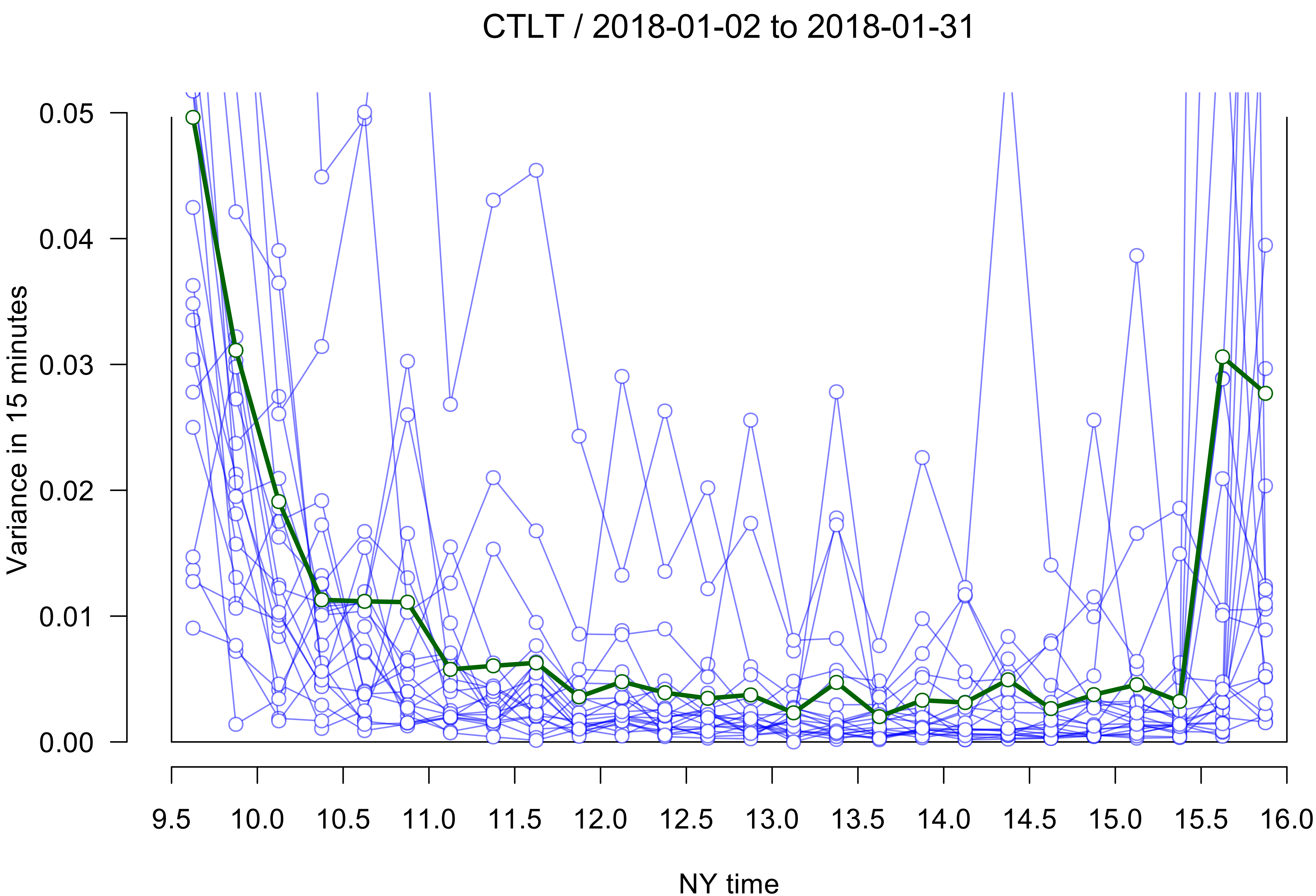
CTLT 2018-01-25
1-minute NBBO returns

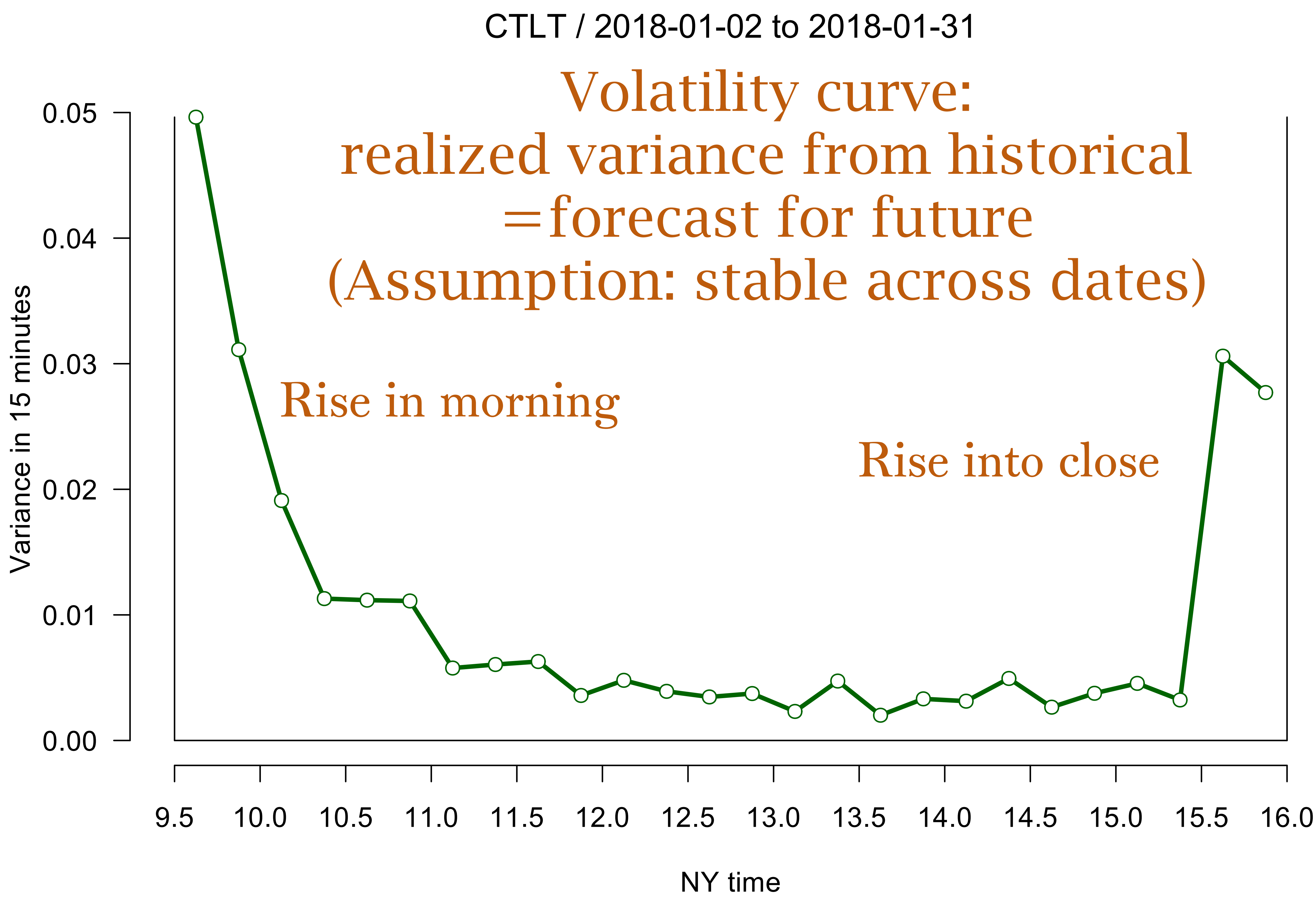


Summary of procedure for realized variance

- Compute NBBO bid and ask at 1-minute time points
  - compute bid and ask from each exchange at each time
  - NBBO bid is max of bids, NBBO ask is min of asks
- Take differences of successive prices
- Take sum of squares of differences in each 15 min
- Compute intraday volatility curves
  by averaging each bin across many days


CTLT / 2018-01-02 to 2018-01-31
Variance in 15 minutes
NY time
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
2018-01-23
2018-01-02
2018-01-03
2018-01-25


CTLT / 2018-01-02 to 2018-01-31
Variance in 15 minutes
0.00
0.01
0.02
0.03
0.04
0.05
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
NY time


CTLT / 2018-01-02 to 2018-01-31
Volatility curve:
realized variance from historical
=forecast for future
(Assumption: stable across dates)
Rise in morning
Rise into close
0.05
0.04
0.03
0.02
0.01
0.00
Variance in 15 minutes
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
NY time

Subtleties to consider later

- Can we use finer time resolution,
  for example to get 1-minute variance?

  yes, but need to account for serial dependence
- Can we detect real-time changes in volatility?

  yes, though not instantly
- Is it stable across months
  in addition to intraday variation?

  more complicated, need product model