



The Global Standard for Business Research

<http://wrds.wharton.upenn.edu>

WRDS E-Learning Session

Overview of TAQ on WRDS

Insights on Efficient Programming in Intraday Research

Denys Glushkov

Mark Keintz

Rabih Moussawi

Luis Palacios

WRDS Research, February 24 2011

E-Learning Session Guidelines

Welcome to WRDS E-Learning. Please use the following guidelines:

- Ask Questions via the “Chat” Window.
- Audio problems? Disconnect and call in using one of the phone numbers listed in the reminder email.

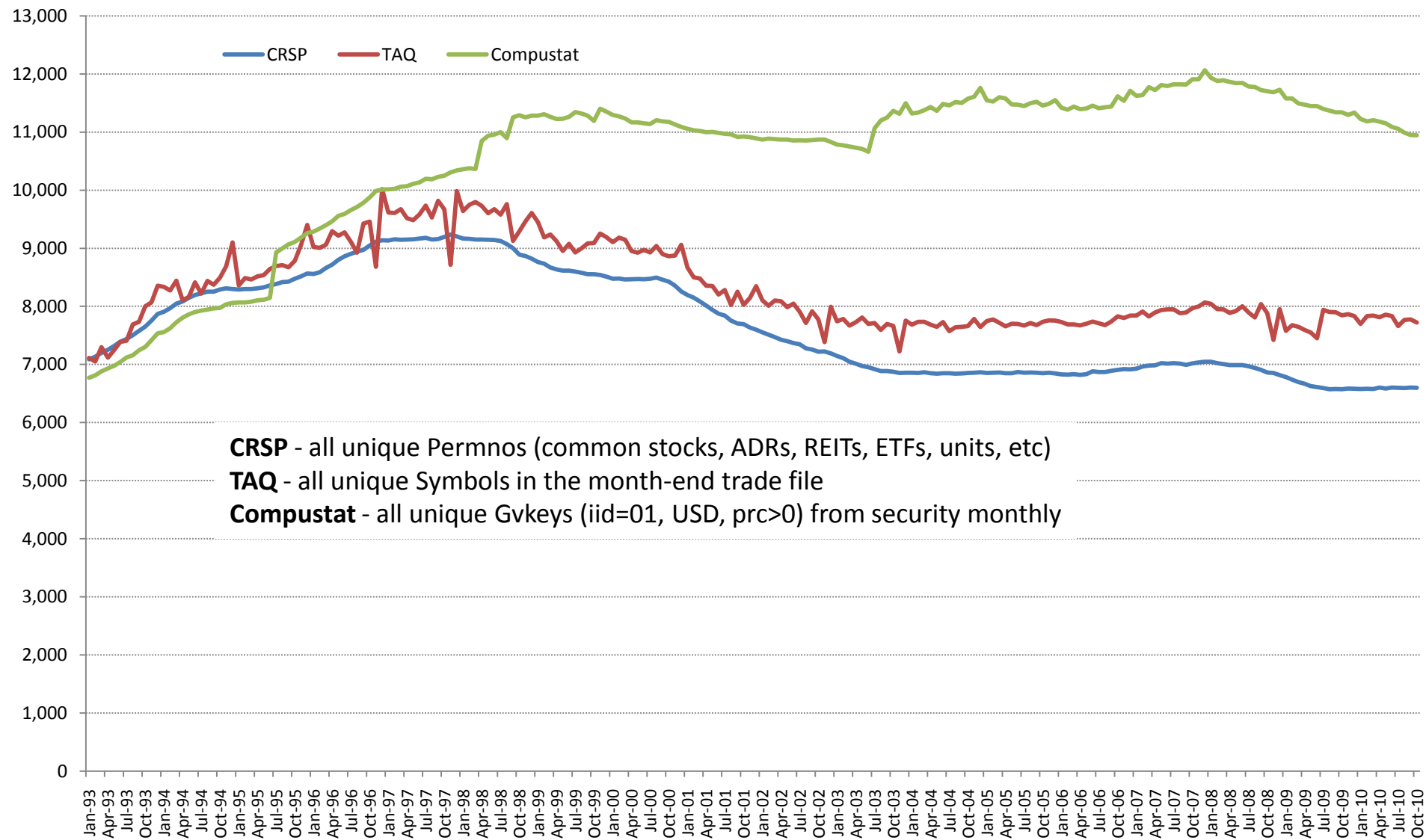
Agenda

- Introduction to TAQ and its research applications
- Efficient Programming with TAQ datasets
- The Lee-Ready Algorithm
- Generating NBBO with TAQ Quotes

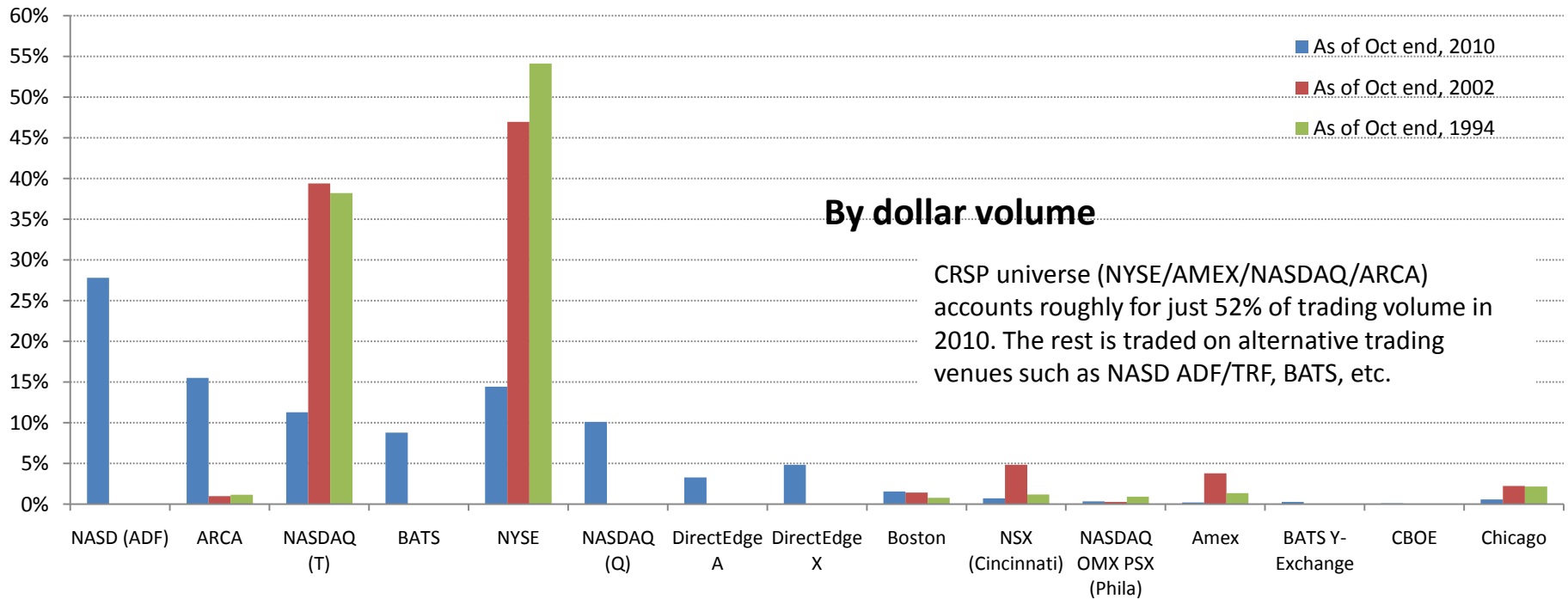
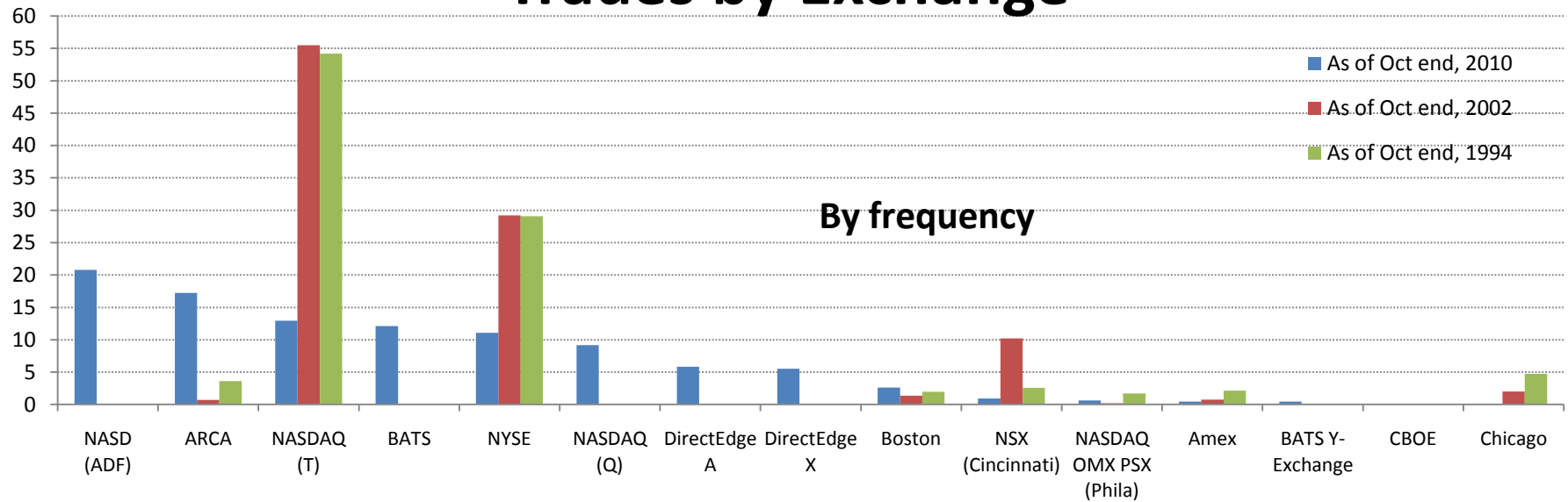
What is TAQ?

- ❑ Trades and Quotes Database contains intraday tick-by-tick transactions data (trades and quotes) for all securities (not aggregate indexes) listed on the NYSE and AMEX as well as Nasdaq National Market System (NMS) and SmallCap issues back to 1993
 - ❑ Excludes transactions reported outside of the Consolidated Tape hours of operation (8.00 am-6.30 pm)
 - ❑ Excludes trading in NYSE-listed securities between 8-9.30 am by other markets
- ❑ As of the latest available vintage (Oct, 2010) TAQ covers more than 7,700 securities traded on 15 different trading venues. The package includes Consolidated Trades& Quotes (time-stamped up to a whole second), Master and Dividend datasets
 - ❑ Common stock -81%, Derivatively priced (e.g., ETFs) – 15.3%, Preferred – 3.3%, Warrant-0.5%, Right-0.03%, other- 0.02% (oct 2010).
 - ❑ TAQ does NOT carry intraday data for either derivatives or OTC securities. OPRA database provides intraday last sale information and options quotations from a committee of Participant Exchanges, whereas Compustat Security file contains daily data for OTC stocks, among others.
- ❑ Size of trades, and especially quotes, datasets expanded exponentially during the last few years, to a large extent due to the proliferation of algorithmic trading
 - ❑ The use of algorithmic trading has grown considerably and, as of 2007, was estimated to account for one-third of equity trading volume in the United States (Hendershott, Jones, and Menkveld, 2010)
 - ❑ A handful of HFT firms (2 percent of the 20,000 trading firms in the US) accounted for an estimated 70% of overall trading volume on US equities markets in 2009 (WSJ, Jul 9, 2009)

Number of distinct traded securities: TAQ (end of month) vs. CRSP vs. Compustat, Jan 1993-Oct 2010



Trades by Exchange



Trading Venues for US equities, Jan-Jun 2008

ECN- Electronic Communication Network

ATS-Alternative Trading System

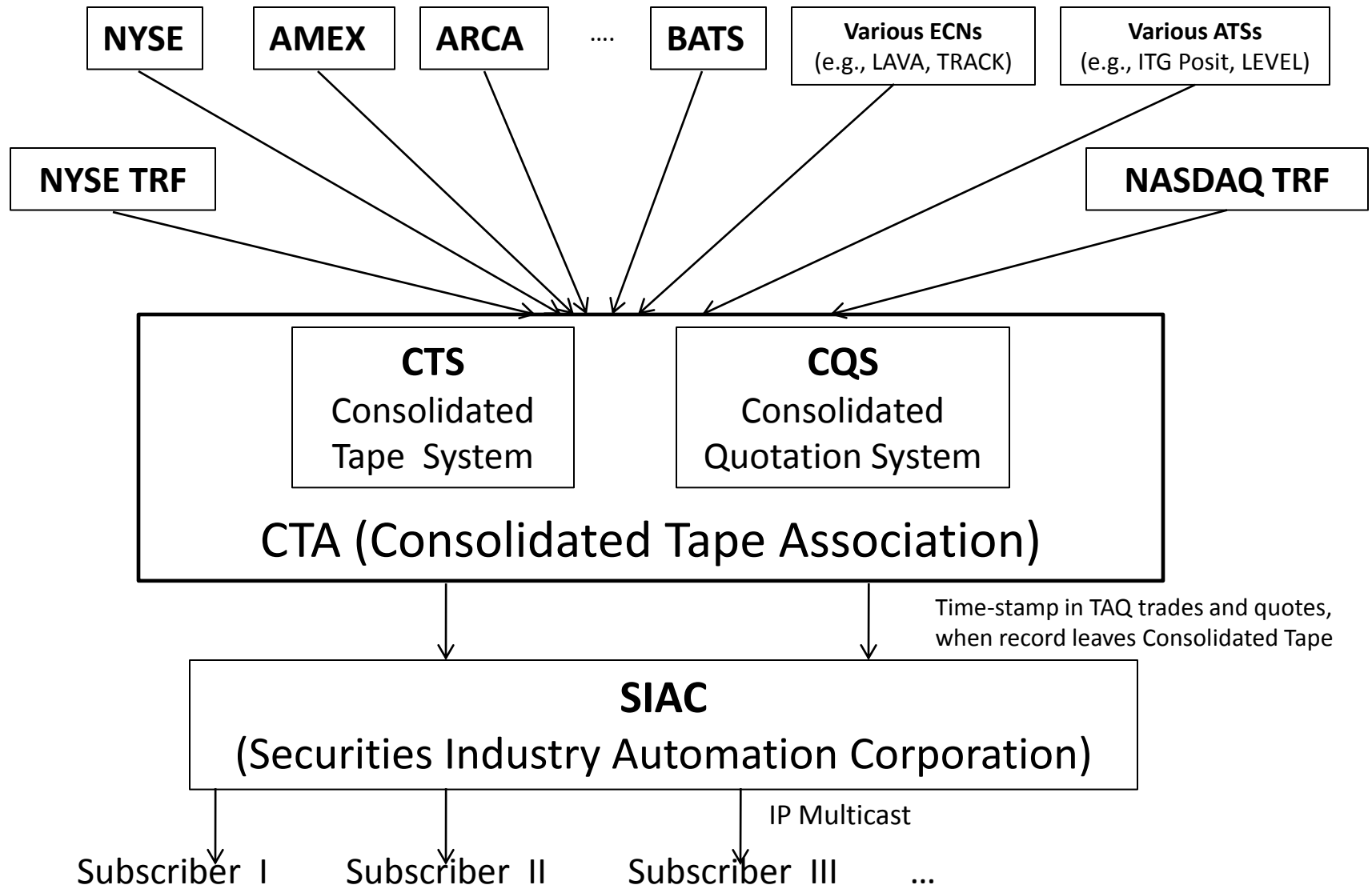
EXCHANGES	ECNS	ATS	
NASDAQ	BATS	ITG POSIT	CITIMATCH
NEW YORK STOCK EXCHANGE	DIRECTEDGE	BIDS	CS CROSSFINDER
ARCA PELAGO	TRADEBOOK	LEVEL	LX
NATIONAL STOCK EXCHANGE	LAVA	LIQUIDNET	MLXN
AMERICAN STOCK EXCHANGE	TRACK	MATCHPOINT	SIGMA X
CHICAGO STOCK EXCHANGE		INSTINCT	MORGAN STANLEY POOL
PHILADELPHIA STOCK EXCHANGE		MILLENNIUM	UBS PIN
BOSTON STOCK EXCHANGE		PIPELINE	BNY CONVERGEX
INTERNATIONAL STOCK EXCHANGE		PULSE	FIDELITY CROSS STREAM
CHICAGO BOARD OPTIONS EXCHANGE		ESPEED AQUA	LAVA ATS

Source: O'Hara and Ye (2010), period Jan-Jun 2008

Regulation (Reg) NMS and Trade Reporting Facilities (TRFs)

- Reg NMS (2005, <http://www.sec.gov/rules/final/34-51808.pdf>) required traders to transact on a trading venue at the lowest price rather than on a venue offering the quickest execution or the most reliability. Imposed caps on access charges that exchanges and other venues could levy - > explosion of trading venues (>40 platforms in 2008)
- Before the SEC mandate off-exchange volume was aggregated with exchange-executed volume for reporting purposes. After March 5, 2007 exchanges must report only their on-exchange volumes, with off-exchange volumes handled by TRFs that report directly to the consolidated tape
 - TRF data in TAQ does not disaggregate trades into specific execution venues. So one can only determine the aggregate off-exchange volume per security, rather than specific volume of trading in each of the many non-exchange venues.
- Currently, the NYSE TRF, the Nasdaq TRF, and the National Stock Exchange (NSX) TRF are active. In addition, the Alternative Trade Facility (ADF) also operated as a TRF.
 - The ADF was originally created by the NASD in response to the Nasdaq market's conversion to for-profit status. The ADF includes both a reporting and display facility, allowing trading platforms who do not wish to post quotes on the Nasdaq an alternative venue in which to display quote and trade information (O'Hara and Ye, 2010)

Overview of Trade and Quote Reporting Procedures



Research applications using intraday data

- Signing the order flow to identify buyer- and seller-initiated trades
 - 5-second rule (Lee and Ready, 1991)
 - Revised algorithms for matching trades and quotes (Vergote, 2005; Rosenthal, 2008)
 - Issues with Lee and Ready when identifying short sales (Asquith, Oman and Safaya, 2010; Chakrabarty et al., 2010)
- Estimation of Transaction Costs, Liquidity and Adverse Selection
 - Validation of Bayesian Gibbs estimates of effective cost using TAQ (Hasbrouck, 2010)
 - Various versions of Bid-Ask spread: quoted, effective and realized spreads and price impact (Goeynko, Holden and Trzcinka, 2009; Brennan et al., 2009), high-low spread estimator (Corwin and Schultz, 2010), VWAP (Hu, 2009)
- Investor Behavior
 - Institutional trading (e.g., Campbell, Ramodara and Schwartz (JFE, 2009) – infer daily institutional trading behavior from TAQ using a sophisticated method that best predicts 13-F data from trades of different size)
 - Retail trading (e.g., Barber, Odean and Zhu, 2006; small trades of Hvidkjaer, 2008)

Research applications using intraday data (cont'd)

- Measures of uncertainty and information risk
 - PIN (Easley, Hvidkjaer, O'Hara, 2002)
 - Flow Toxicity/VPIN (Easley, Lopez de Prado, O'Hara, 2010, 2011)
- Impact of decimalization (before and after)
 - Cost of trading and market quality (Bessembinder, 2003)
 - Stock return implications (Graham, Michaely and Roberts, 2003)
- Tests of market efficiency
 - Speed of price discovery, reaction to order flows and imbalances (Chordia, Roll and Subrahmanyam, 2005&2008)
 - Predictable periodicity in intraday returns (Heston, Korajczyk, Sadka, 2010)
- High-Frequency and Algorithmic Trading
 - Impact of HFT on market quality (Brogaard, 2010), Low-latency (Hasbrouck et al., 2011)
 - Floor trader vs automation (Venkataraman, 2001; Cardella, Hao, Kalcheva, 2010)
 - Transparency of the trading process (Bessembinder, Panayides and Venkataraman, 2009)

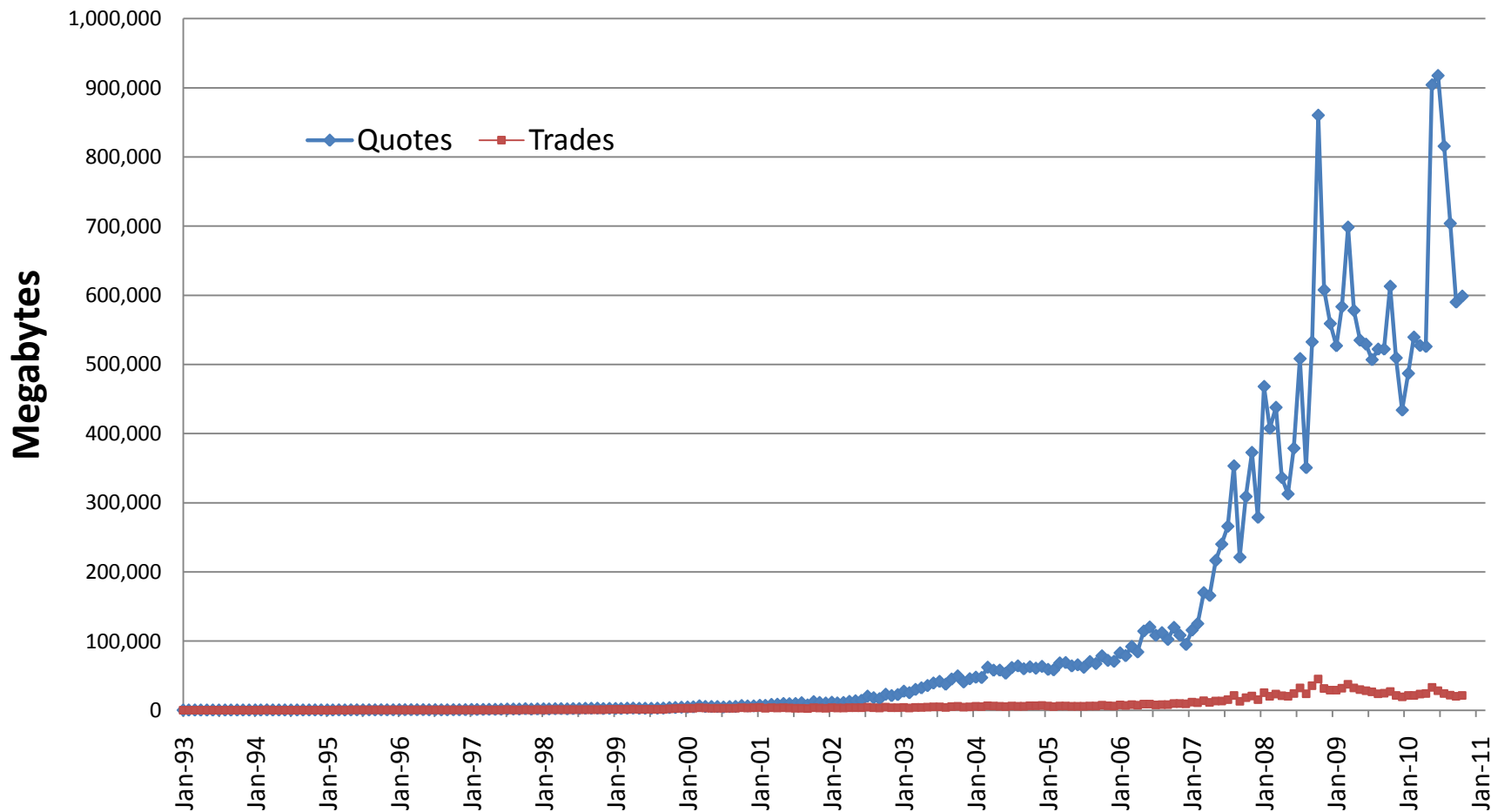
Efficient Programming with the TAQ Database

TAQ Data Organization at WRDS

- Stored as SAS datasets
- Located in `/wrds/taq/sasdata`
- Covers all trades & quotes since Jan 1993
- Occupies about 32 Terabytes
- Typical recent month adds 850 gigabytes

Major Feature of TAQ Data: Size

TAQ Monthly File Size - Jan 1993 to Oct 2010



Four Types of TAQ Data Files

- **CQ: Consolidated Quotes**
 - One record per quote (468 billion for Jan 1993-Oct 2010)
- **CT: Consolidated Trades**
 - One record per trade (39 billion for Jan 1993-Oct 2010)
- **MAST: Master**
 - Tracks the CUSIP and Exchanges for each SYMBOL
- **DIV: Dividends**
 - Provides Dividends by Date and SYMBOL

CT and CQ datasets

“Consolidated” Trades and Quotes

- “Consolidated”
Refers to centralization of record processing
NOT to aggregation of trades or quotes.
- One dataset per trading date.
- Named as CX_yyyymmdd:
CT_19930104 covers trades on Jan 4, 1993
CQ_20101001 covers quotes on Oct 1, 2010

CT and CQ datasets

Large Size and Tied TIME Values

- Each CQ/CT dataset indexed by SYMBOL
 - Support faster filtering by SYMBOL
- Each dataset sorted by SYMBOL TIME
 - Supports “BY processing” (see WRDS SAS e-learning)
- TIME is in whole seconds, yielding many ties:
 - CT files: 5% (1993) to 88% (2010)
 - CQ files: 25% (1993) to 98% (2010)
- Records with tied TIME from data supplier are not reordered in WRDS datasets.

CT Dataset Contents

- One Record per Trade

Sample Contents from CT 20100101

[illegible]

CQ Dataset Contents

- One Record per Quote

Sample Contents from CQ 20100101

[illegible]

Common TAQ Programming Issues

- Large TAQ datasets engender excessive resource use
 - Disk I/O (input/output)
 - Disk storage space for intermediate datasets
 - ***LONG TIME TO COMPLETION***
- The Primary Pitfalls of User Programs are:
 - ***EXCESSIVE USE OF PROC SORT***
 - Inefficient extraction of subsets
 - Generating unneeded intermediate datasets
 - Using PROC SQL to sort TAQ data

Listing Multiple Datasets With One Name Prefix

- The **SET** statement accepts *name prefix*.
(a partial dataset name truncated by a “:”)
- Read all October 2010 trade datasets with:
set taq.ct_201010: ;
- Other examples (don't try these):
set taq.cq_: ; ** All TAQ quotes **;
set taq.ct_2005: ; ** All 2005 trades **;

Sample Data Extraction

```
libname taq '/wrds/taq/sasdata';

data mydata;
  set taq.cq_201010: ;
  where symbol in ("IBM", "DELL", "MSFT", "CSCO");
  if bid>0 and ofr>bid and (ofr-bid)/(ofr+bid) < .1;
  spread=ofr-bid;
run;

proc means data=mydata n min max mean;
  class symbol;
  var spread;
run;
```

This program took 2 Minutes, 16 Seconds

Replace intermediate datasets with Views

Making a data VIEW does not write to disk.

It “pipes” the data directly to the PROC MEANS

```
data v_mydata / view=v_mydata;  
    set taq.cq_201010: ;  
    where symbol in ("IBM","DELL","MSFT","CSCO");  
    if bid>0 and ofr>bid and (ofr-bid)/(ofr+bid) < .1;  
    spread=ofr-bid;  
run;  
proc means data=v_mydata n min max mean;  
    class symbol;  
    var spread;  
run;
```

This program took 1 Minute, 12 Seconds

Time Filters with TAQ Data

You can use “Time Literals” to filter the TIME variable

```
data vmydata / view=vmydata;  
  set taq.cq_20101001;  
  where symbol in ('IBM','DELL')  
    and time between '09:30:00't and '16:00:00't;  
  ... other statements here ...;  
run;  
  
proc means data=vmydata;  
  ...  
run;
```


Sorting by DATE/SYMBOL/TIME IS SUPERFLUOUS

This SET statement reads all datasets with prefix CQ_201010 in alphabetic order (same as date order)

```
data vtemp / view=vtemp;  
  set taq.cq_201010: ;  
  where symbol in ('IBM','DELL');  
run;
```

SET reads all the cq_201010xx datasets consecutively, so VTEMP is already sorted by DATE SYMBOL TIME. As a result ...

THIS SORT IS SUPERFLUOUS

```
proc sort data=vtemp out=temp;  ** NOT NEEDED **;  
  by date symbol time;  
run;
```

“OPEN=DEFER” Saves Memory

Accessing multiple daily datasets (instead of one monthly dataset) uses more memory. SAS assigns a memory buffer for each dataset read by the SET statement.

```
data mydata;  
  set taq.cq_201010: ;  
  where symbol in ('IBM', 'DELL');  
run;
```

Memory Use: 573K

Using “open=defer” tells SAS to “reuse” the same memory buffer for each incoming data set

```
data mydata;  
  set taq.cq_201010: open=defer;  
  where symbol in ('IBM', 'DELL');  
run;
```

Memory Use: 339K

Ordering by SYMBOL/DATE/TIME

The Naïve Sort

This SET reads datasets in order, yielding DATE/SYMBOL/TIME order.

```
data vtemp /view=vtemp;  
  set taq.cq_201010: ;  
  where symbol in ('IBM','DELL');  
run;
```

To get SYMBOL/DATE/TIME order, a PROC SORT is needed

```
proc sort data=vtemp out=mydata;  
  by symbol date time;  
run;
```

Note: In PROC SORT, records with tied TIME values are NOT reordered

Ordering by SYMBOL/DATE/TIME

Avoiding the Naïve Sort

Using a BY statement with the SET statement will allow the incoming data to be “interleaved”. I.e. all the “A” SYMBOLs are processed first, then the “B” SYMBOLs, etc.

```
data mydata;  
  set taq.cq_201010: ;  
  by symbol date time;  
  where symbol in ('IBM', 'DELL');  
run;
```

Because the incoming datasets have a constant DATE value, and are sorted by SYMBOL/TIME, the datasets can be read “in parallel” to produce the desired order, *without using PROC SORT*.

PROC SORT vs. “order by” (PROC SQL)

- PROC SORT guarantees order preservation for tied records.
- The PROC SQL “order by” clause does not guarantee order preservation for tied records (e.g. “order by symbol,date,time”)
- WRDS Strongly Recommends using PROC SORT for ordering TAQ data rather than PROC SQL

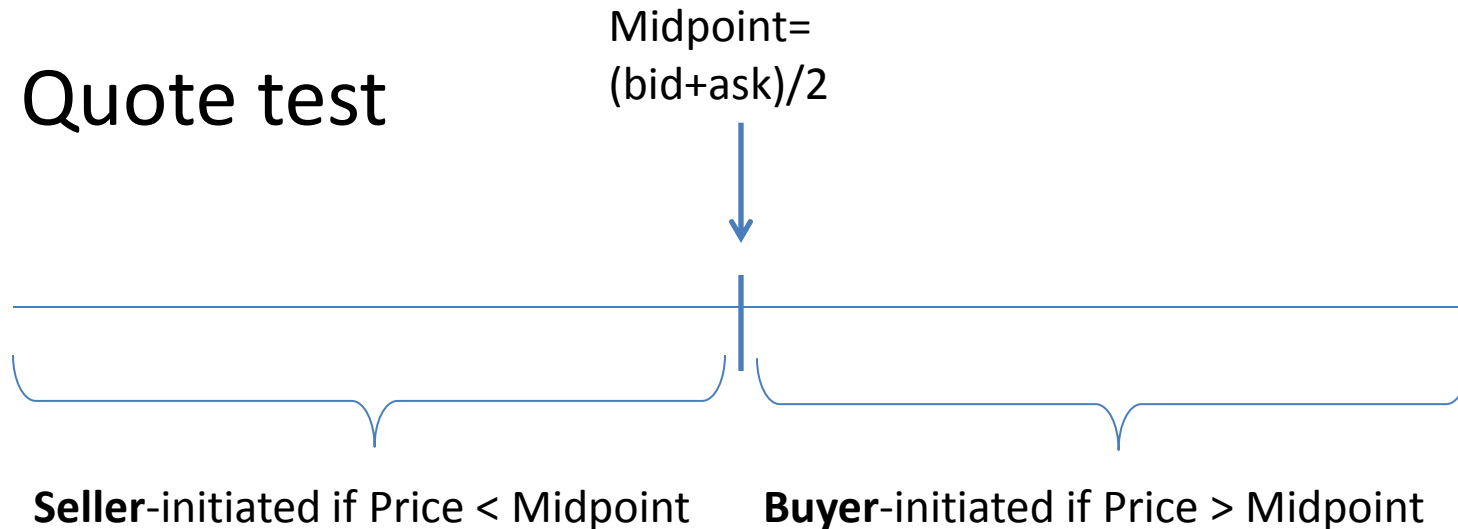
SAS and TAQ Links at WRDS

- SAS E-learning links on:
<http://wrds-web.wharton.upenn.edu/wrds/E-Learning/>
"SAS at WRDS"
"Intermediate SAS at WRDS"
- Other comments on Using TAQ Efficiently:
http://wrds-web.wharton.upenn.edu/wrds/support/Data/_005Using%20TAQ%20Data%20Efficiently/index.cfm
- PROC SQL "ORDER BY" with tied records:
[http://wrds-web.wharton.upenn.edu/wrds/support/Data/_005Using%20TAQ%20Data%20Efficiently/_017Note on Usage of SQL with TAQ data.cfm](http://wrds-web.wharton.upenn.edu/wrds/support/Data/_005Using%20TAQ%20Data%20Efficiently/_017Note%20on%20Usage%20of%20SQL%20with%20TAQ%20data.cfm)

Lee and Ready (1991) Algorithm

Who initiated a trade?

1. Quote test



2. Tick test if price=midpoint

-**Seller**-initiated: $P(t) < P(t-1)$

-**Buyer**-initiated : $P(t) > P(t-1)$

(in case of equality, the procedure can go back several lags).

Are 'Trades' and 'Quotes' aligned ?

Delay of the reported trade time wrt a quote time

- There are several assumptions about misalignment between trade time and reported quote time
- Lee-Ready (1991), Peterson and Sirri (2003), Henker and Wang (2005)
- Lee-Ready for example use a “5-second” difference between trade time and quote time.

WRDS-sample program

Main Characteristics and Efficiency Issues

- Lee and Ready Sample SAS program:
http://wrds-web.wharton.upenn.edu/wrds/support/Data/004Research%20Applications/010Market%20Microstructure%20Study/lee_ready.cfm
- Use SAS Views.
- Use 5-second rule (can be easily modified).
- Multiple entries by second: assumptions needed. Currently, no millisecond time variable=> how to merge quotes and trades?
- Take advantage that TAQ files are sorted by SYMBOL and TIME.

```
data lee_ready;  
  set   quote_file   trade_file;  
  by    date   symbol   time;  
  ...
```

Output-sample

Stock Symbol	Quote date	Quote time	Price	mid_point_Bid_Ask	lagprice	lag2price	TICK Test	buyer_init	
DELL	20090120	9:45:01	10.177	10.185	10.18	10.19	-1	-1	seller
DELL	20090120	9:45:29	10.164	10.160	10.16	10.1607	1	1	buyer
DELL	20090120	9:46:47	10.125	10.125	10.13	10.1254	-1	-1	seller
DELL	20090120	9:48:23	10.085	10.085	10.08	10.0804	1	1	buyer
DELL	20090120	9:49:24	10.08	10.08	10.08	10.0888	-1	-1	seller
DELL	20090120	9:54:39	10.02	10.02	10.02	10.02	0	0	not defined

Generating NBBO From TAQ Quotes

NBBO Research Application

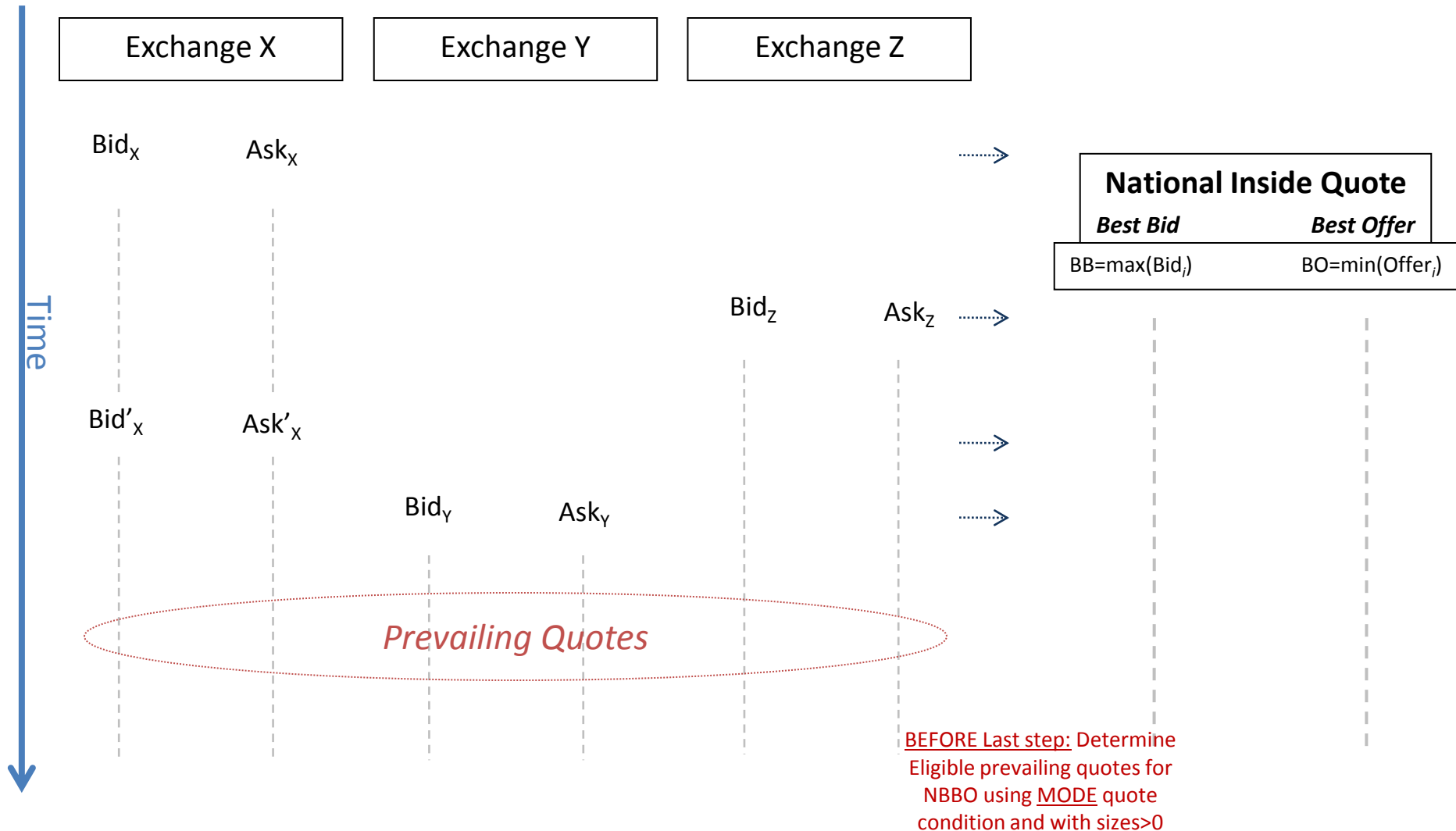
- Regulation NMS – National Market System: calculate **National** Inside Quote from valid quotes across various exchanges: Best Bid (BB), Best Offer (BO), Size at BB, Size at BO
- Quote Rule: <http://www.sec.gov/answers/trdexbd.htm>
 - best price at which a market maker is willing to trade
 - a prevailing quote of an exchange is the most recent valid quote from that exchange
 - every new quote by an exchange supersedes its previous prevailing quote
- NBBO calculation every second because of TAQ Monthly feed time stamp granularity
- Evaluation of accuracy of WRDS NBBO algorithm:
http://wrds-web.wharton.upenn.edu/wrds/support/Data/_005Using%20TAQ%20Data%20Efficiently/_028NBBO_note.cfm
 - *Minimize* the percentage of *locks* and *crosses* within an NBBO
 - *Maximize* the percentage of *Inside Trades*
- SAS Code: http://wrds-web.wharton.upenn.edu/wrds/support/Data/_005Using%20TAQ%20Data%20Efficiently/_028NBBO.cfm
- Invocation: `%nbbo (yyyyymmdd = 20080606 , outset = myfile) ;`
- Evaluation of execution speed of WRDS NBBO algorithm SAS Macro on WRDS server
 - To create a 1998 NBBO daily dataset for all Stocks: 26 seconds
 - To create a 2006 NBBO daily dataset for all Stocks: 9 minutes 37 seconds

NBBO Algorithm: TAQ Quote Tape



Stock Symbol	Time	Bid	Ask	Exchange	Mode (quote condition)
ABC	11:30:02	Bid _x (1)	Ask _x (1)	X	12
ABC	11:30:05	Bid _z (2)	Ask _z (2)	Z	12
ABC	11:30:06	Bid _y (3)	Ask _y (3)	Y	12
ABC	11:30:08	Bid _x (4)	Ask _x (4)	X	
ABC	11:30:40	Bid _y (5)	Ask _y (5)	Y	10
ABC	11:31:22	Bid _y (6)	Ask _y (6)	Y	12
ABC	11:32:50	Bid _z (7)	Ask _z (7)	Z	12

NBBO Algorithm: SAS Code Logic



NBBO Output Example: IBM

Stock Symbol	Quote date	Quote time	Best Bid	Best Offer	Best Bid Size	Best Offer Size	Total Number of New Quotes	# of Exchanges with Valid Prevailing Quotes
IBM	19970303	9:47:13	144.625	144.875	36	513	1	7
IBM	19970303	9:47:16	144.625	144.875	36	508	1	7
IBM	19970303	9:47:34	144.75	144.875	50	508	1	7
IBM	19970303	9:47:35	144.75	144.875	50	508	1	7
IBM	19970303	9:47:36	144.75	144.875	51	508	1	7
IBM	19970303	9:47:37	144.75	144.875	51	508	2	7
IBM	19970303	9:47:38	144.75	144.875	51	508	2	7
IBM	19970303	9:47:39	144.75	144.875	101	513	3	7
IBM	19970303	9:47:40	144.75	144.875	101	513	3	7
IBM	19970303	9:47:45	144.75	144.875	101	513	1	7
IBM	19970303	9:48:07	144.75	144.875	151	513	1	7
IBM	19970303	9:48:20	144.75	144.875	151	613	1	7
IBM	19970303	9:48:27	144.75	144.875	166	608	1	7
IBM	19970303	9:48:56	144.75	144.875	16	608	1	7
IBM	19970303	9:48:58	144.75	144.875	15	608	1	7
IBM	19970303	9:49:00	144.75	144.875	15	608	4	7
IBM	19970303	9:49:01	144.75	144.875	10	608	1	7
IBM	19970303	9:49:05	144.625	144.875	60	613	2	7
IBM	19970303	9:49:06	144.75	144.875	50	613	1	7

Future of Microstructure Research on WRDS

- WRDS High-Performance TAQ Platform (WRDS HPP)
 - The new WRDS HPP is optimized for TAQ access while the current WRDS server is a general-use server
 - Users on WRDS HPP will be able to process ALL TAQ data in days when it could take weeks otherwise
 - The new WRDS HPP contains TAQ millisecond data, as well as TAQ monthly feed data, and will include TAQ-CRSP linking tables, and NBBO datasets in SAS format
 - WRDS High-Performance TAQ Platform is available at an additional cost. Please contact WRDS-Support@wharton.upenn.edu for more information
- Research Applications Work-in-Progress
 - TAQ-CRSP Linking Macro: TCLINK.SAS that maps PERMNO to TAQ Symbol
 - Buy-Sell Imbalance Research Application
 - Intraday Return and Intraday Event Study Research Application

References

1. Asquith, P., Oman, R. and Safaya, C., 2008. "Short Sales and Trade Classification Algorithms", working paper, MIT
2. Barber, B.M., Odean, T. and Zhu, N., 2006. "Do noise traders move markets?", working paper, UC Davis
3. Bessembinder, H., 2003. "Quote-based competition and trade execution costs in NYSE-listed stocks", *Journal of Financial Economics*, Vol. 70, Issue 3
4. Bessembinder, H., Panayides, M. and Venkataraman, K., 2009. "Hidden Liquidity: an Analysis of order exposure strategies in electronic stock markets", *Journal of Financial Economics*, Vol. 94, Issue 3, Dec
5. Brennan M., Chordia, T., Subrahmanyam, A., and Tong, Q., 2009. "Sell-Side Illiquidity and the Cross-Section of Expected Stock Returns", working paper, UCLA
6. Brogaard, J., 2010. "High Frequency Trading and its Impact on Market Quality", working paper, Northwestern University
7. Campbell, J.Y., Ramadorai, T. and Schwartz, A., "Caught on Tape: Institutional Trading, stock returns, and earnings announcements", *Journal of Financial Economics*, vol. 92, issues 1, April.
8. Cardella, L., Hao, J. and Kalcheva, I., 2010. "The Floor Trader vs. Automation: A survey of theory and empirical evidence", working paper, University of Arizona
9. Chakrabarty, B., Moulton, P.C. and Shkilko, A., 2010. "Short-Sales and the Lee-Ready Trade Classification Algorithm Revisited", working paper, Saint Louis University
10. Chordia, T., Roll, R. and Subrahmanyam, A., 2005. "Evidence on the speed of convergence to market efficiency", *Journal of Financial Economics*, Vol. 76, Issue 2, May
11. Chordia, T., Roll, R. and Subrahmanyam, A., 2008. "Liquidity and Market Efficiency", *Journal of Financial Economics*, 87.
12. Corwin, S.A. and Schultz, P.H., 2010. "A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices", working paper, University of Notre Dame
13. Easley, D., Hvidkjaer, S., and O'hara, M., 2002. "Is Information Risk a Determinant of Asset Returns?", *Journal of Finance*, Vol. LVII, No.5, Oct
14. Easley, D., Lopez de Prado, M. and O'Hara, M., 2010. "Measuring Flow Toxicity in a High Frequency World", working paper, Cornell University
15. Easley, D., Lopez de Prado, M. and O'Hara, M., 2011. "The Microstructure of the 'Flash Crash': flow toxicity, liquidity crashes and the probability of informed trading", working paper, Cornell University
16. Goyenko, R.Y., Holden, C.W., and Trzcinka, C.A., 2009. "Do liquidity measures measure liquidity?", *Journal of Financial Economics*, 92, pp. 152-181.
17. Graham, J.R., Michaely, R. and Roberts, M.R., 2003. "Do Price Discreteness and Transaction Costs Affect Stock Returns? Comparing Ex-Dividend Pricing before and after Decimalization", *Journal of Finance*, Vol. 58, No.6, Dec
18. Hasbrouck, J., 2010. "Trading Costs and Returns for US equities: Estimating Effective Costs from Daily Data", *Journal of Finance*, forthcoming
19. Hasbrouck, J. and Saar, G., 2011. "Low-latency trading", working paper, New York University
20. Hendershott, T., Jones C.M. and Menkveld, A.J., 2010. "Does Algorithmic Trading Improve Liquidity?", *Journal of Finance*, forthcoming
21. Henker, T. and Wang, J., 2005. "On the Importance of Timing Specifications in Market Microstructure Research", working paper, UNSW, Australia
22. Heston, S.L., Korajczyk, R. and Sadka, R., 2010. "Intraday Patterns in the Cross-Section of Stock Returns", *Journal of Finance*, Vol 65, Issue 4, Aug
23. Hvidkjaer, S., 2008. "Small Trades and the Cross-Section of Stock Returns", *Review of Financial Studies*, 21(3)
24. Hu, Gang, 2009. "VWAP Cost Excluding Own Trades", *Journal of Trading*, Vol. 2, pp.30-34
25. Lee, Charles MC and Ready, M.J., 1991. "Inferring Trade Direction from Interday Data", *Journal of Finance*, vol. 46(2), pp. 733-46, June
26. O'hara, M. and Ye, M., 2009. "Is Market Fragmentation Harming Market Quality?", working paper, Cornell University
27. Peterson, M., and Sirri, E., 2003. "Evaluation of the biases in execution cost estimation using trade and quote data", *Journal of Financial Markets*, Vol. 6, Issue 3
28. Rosenthal, Dale W.R., 2008. "Modeling Trade Direction", working paper, University of Illinois at Chicago
29. Venkataraman, K., 2001. "Automated vs. Floor Trading: an Analysis of Execution Costs on the Paris and New York Exchanges", *Journal of Finance*, Vol. 56, No. 4,
30. Vergote, O., 2005. "How to match Trades and Quotes for NYSE stocks?", working paper, Catholic University Leuven
31. Weidner, David, 2009. "Subplots of the Goldman code heist", MarketBeat blog, July 9, <http://blogs.wsj.com/marketbeat/2009/07/09/subplots-of-the-goldman-code-heist/>