

Filtering High Frequency Data

SIA 2004 Technology Management Conference
June 10, 2004

Neal Falkenberry, CFA
CEO
Tick Data, Inc.



Arthur Whitney
CTO
Kx Systems



Presented By

Neal Falkenberry, CFA – CEO, Tick Data, Inc.

Mr. Falkenberry is the majority owner and CEO of Tick Data, Inc. In addition, he founded and serves as Chief Investment Officer for Autumn Wind Asset Management, an SEC-registered investment advisory firm and the General Partner to Autumn Wind Capital Partners, LLC, a commodity pool operator. Earlier, Mr. Falkenberry spent 15 years as an institutional portfolio manager with First Union's West Palm Beach, FL and Washington, DC offices.

Arthur Whitney - CTO, Kx Systems

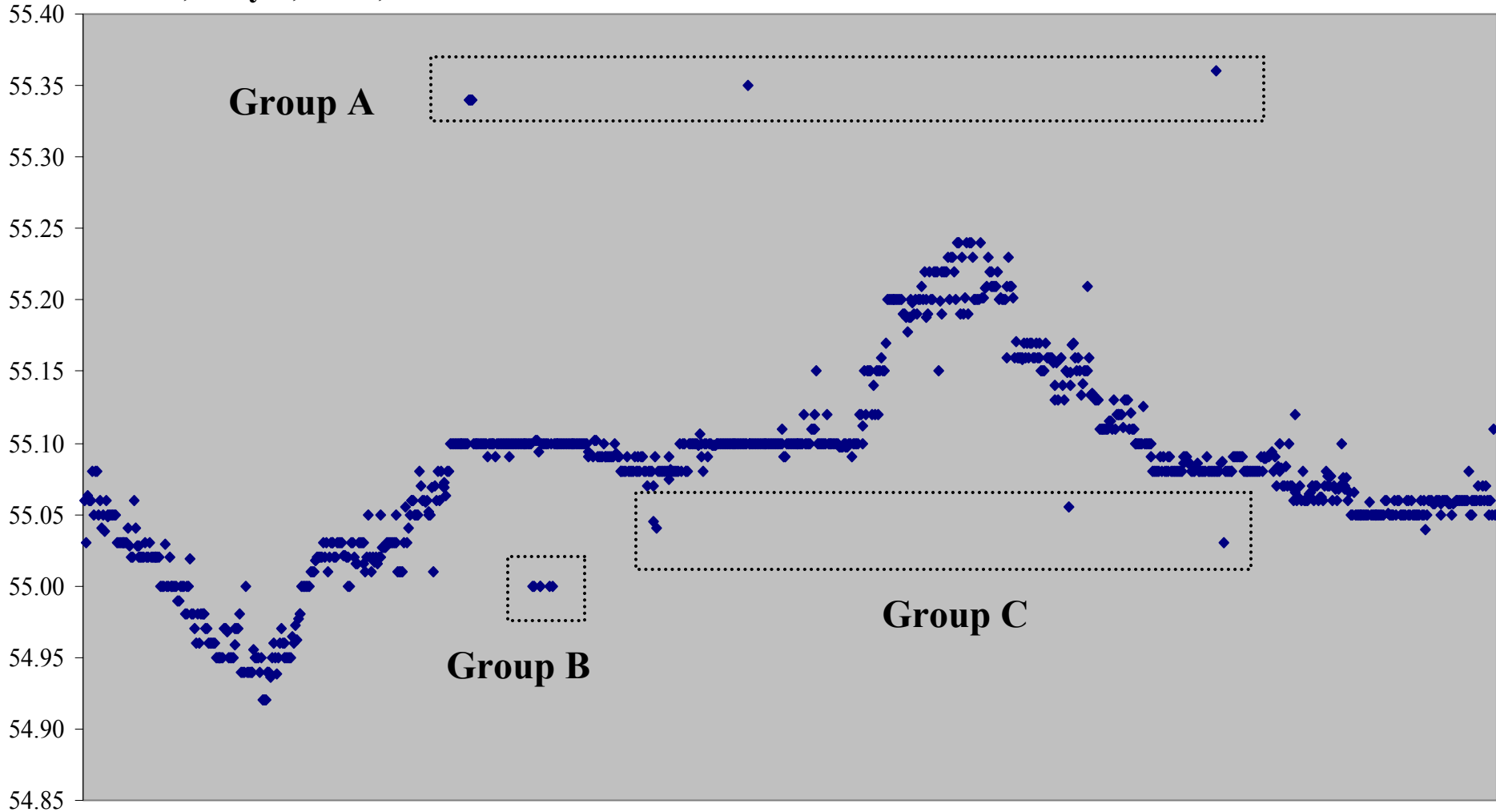
Prior to founding Kx, Mr. Whitney was a Managing Director of Union Bank of Switzerland (UBS) in New York, where he led an internal team that developed global trading and risk management systems using the k language. Earlier, Mr. Whitney was at Morgan Stanley, where he developed the A+ programming language, used to build trading systems, databases and analytics for equities and fixed income. Mr. Whitney studied set theory, foundations and computational complexity at the University of Toronto and Stanford

Topics

1. Define the Problem of “High Frequency Data Filtering”
 1. What is “bad”?
 2. How common is bad data?
2. Why is it important?
 1. Validity of backtesting.
 2. Electronic execution
3. What are the issues?
4. What are the solutions?
5. Market Data
6. Continuous Queries
7. Corporate actions and indices

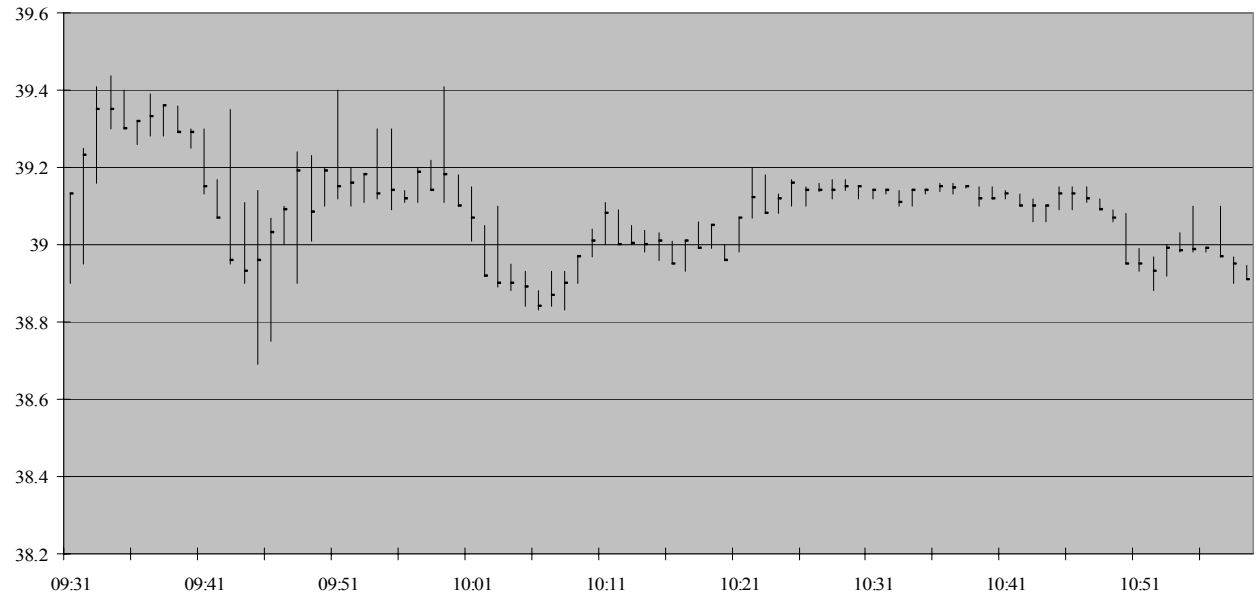
The Problem

MSFT, May 2, 2002, 9:40-9:43

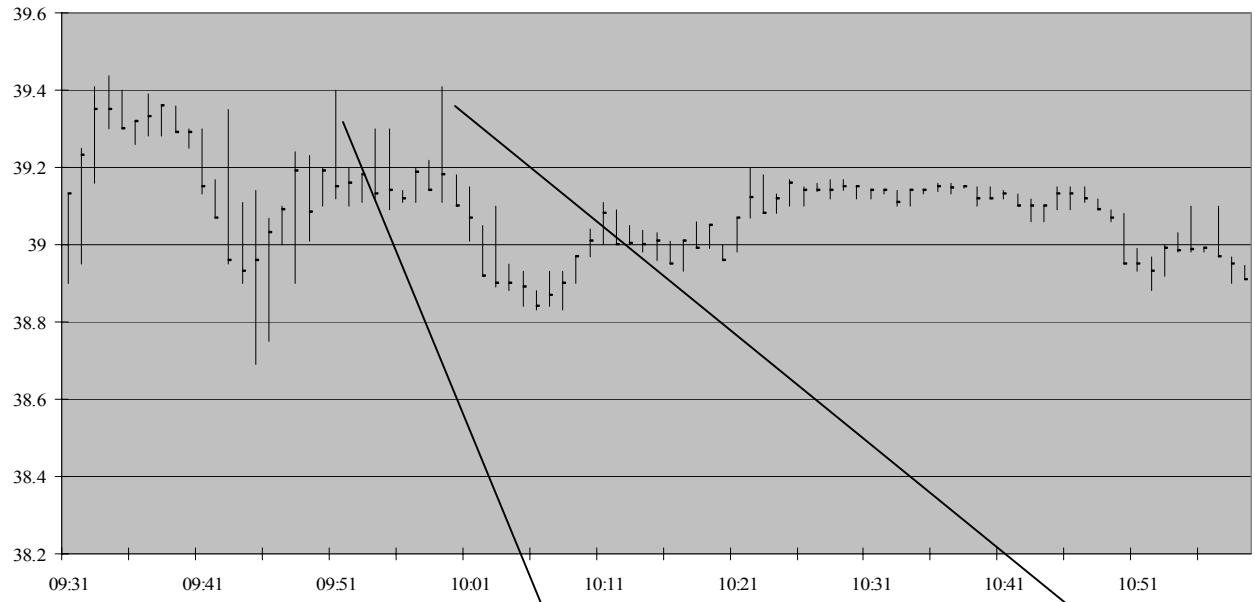


COST, May 30, 2002, 9:30-11:00

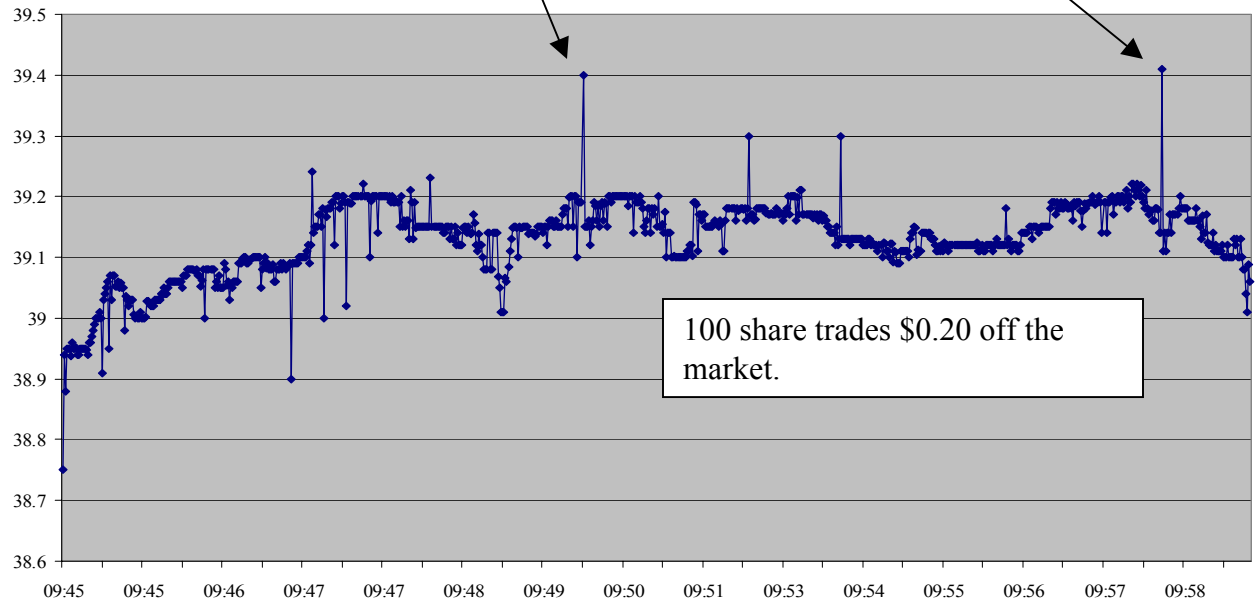
**What appears as
“clean” data to a trader
using 1-minute bars**



COST, May 30, 2002, 9:30-11:00

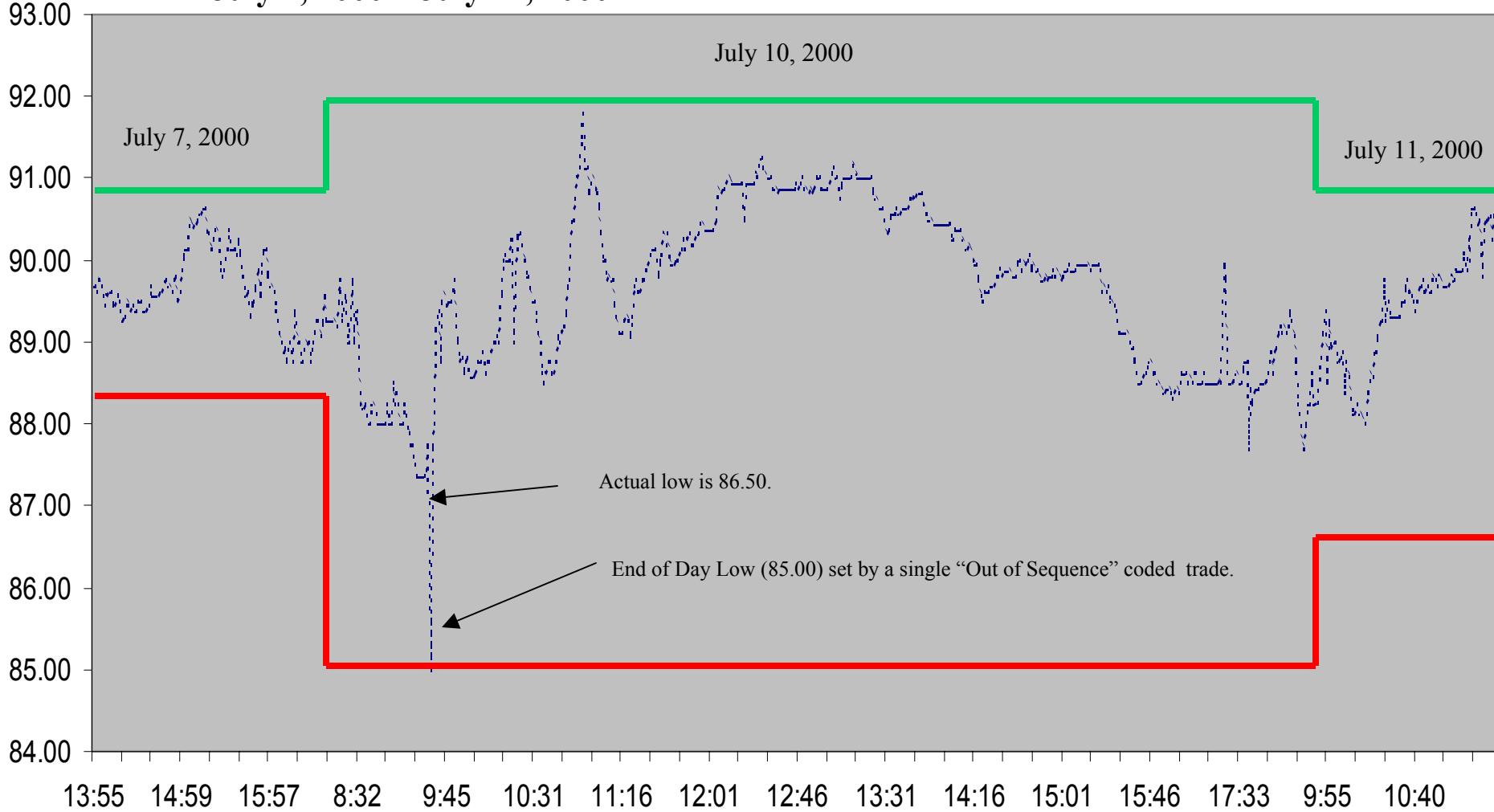


Is “unclean” to a trader using tick data.



Problem Extends to End-of-Day Data

AMAT July 7, 2000 – July 11, 2000



Tick Error Rates

Trade Data: January 1, 2004 – March 31, 2004

	AMEX		NASDAQ		NYSE		ALL
	SPY	DVN	BMET	AMGN	GE	ATH	
I. Total Ticks							
Trade Data	1,955,041	106,708	387,058	1,491,852	875,089	171,761	4,987,509
II. Valid Ticks							
No Exceptions	1,935,932	105,166	383,549	1,482,488	860,991	170,824	4,938,950
III. Bad Ticks - Condition Code Issued							
Condition Code	17,800	1,173	3,303	8,572	12,831	853	44,532
IV. Bad Ticks - No Condition Code Issued							
Algorithmic	1,309	369	206	792	1,267	84	4,027
Total "Bad" Ticks	19,109	1,542	3,509	9,364	14,098	937	48,559
Total "Bad" Ticks (%)	0.98%	1.45%	0.91%	0.63%	1.61%	0.55%	
Avg Ticks per Day	31,533	1,721	6,243	24,062	14,114	2,770	
Avg Bad Ticks per Day	308	25	57	151	227	15	
Minutes between Bad Ticks	1.3	16.3	7.2	2.7	1.8	26.8	

Key Points and Objective

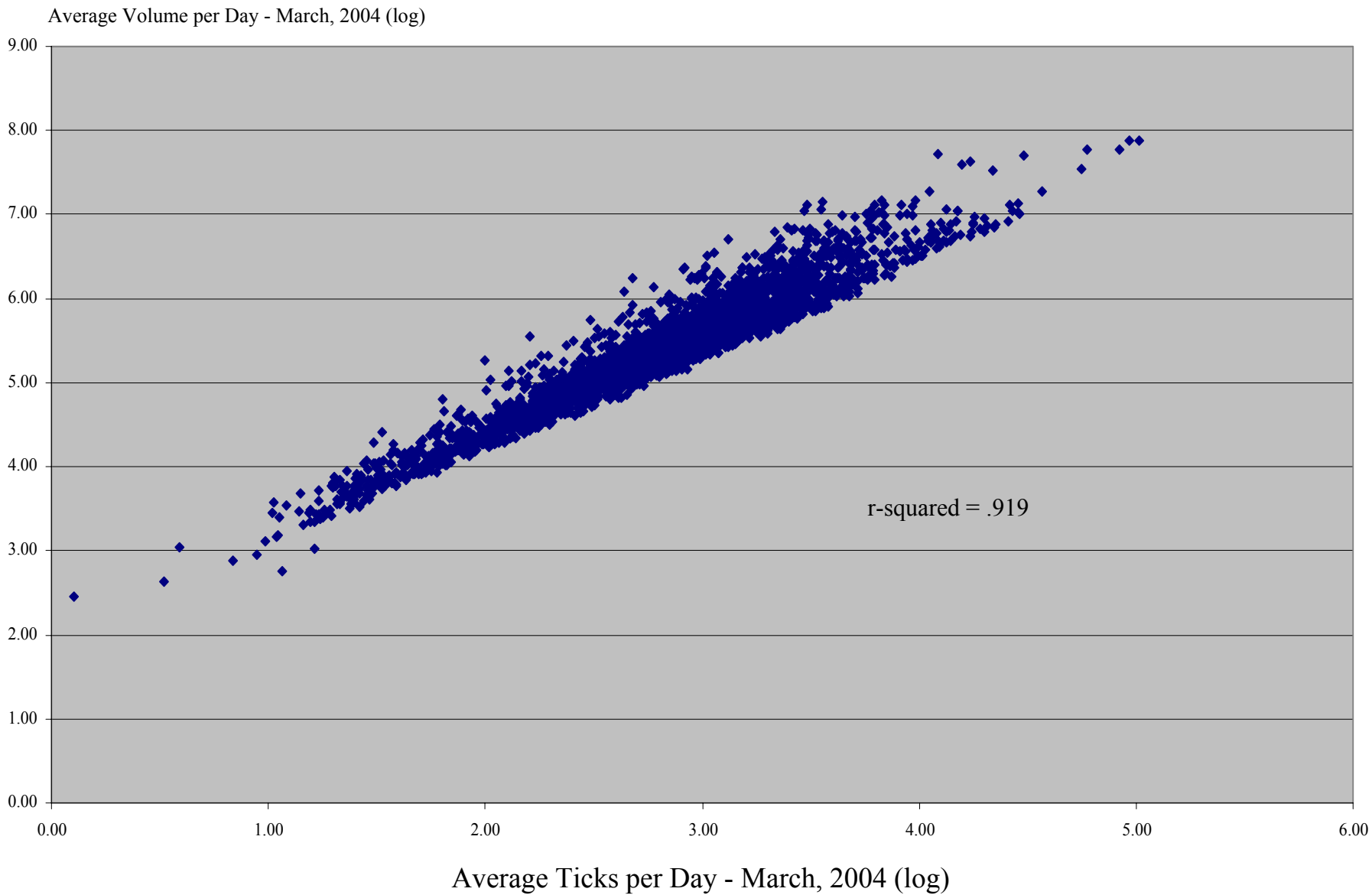
- Data must be filtered at its finest granularity.
- Overscrubbing data is a serious form of risk.
- Do not clean historical data in a fashion that cannot be replicated in real-time.
- The definition of “*bad*” tick is unique to each trader and is a function of the trader’s time scale, trading strategy, and what is possible in real-time. There is no single “correct” filtered data set.
- Match historical and real-time data practices. Do not change the micro-structural properties of historical data vis-à-vis real-time.

The primary objective in developing a set of tick filters is to manage the overscrub/underscrub tradeoff in such a fashion as to produce a time series that removes outliers in the trader’s base unit of analysis so as to support legitimate historical backtesting without removing real-time properties of the data.

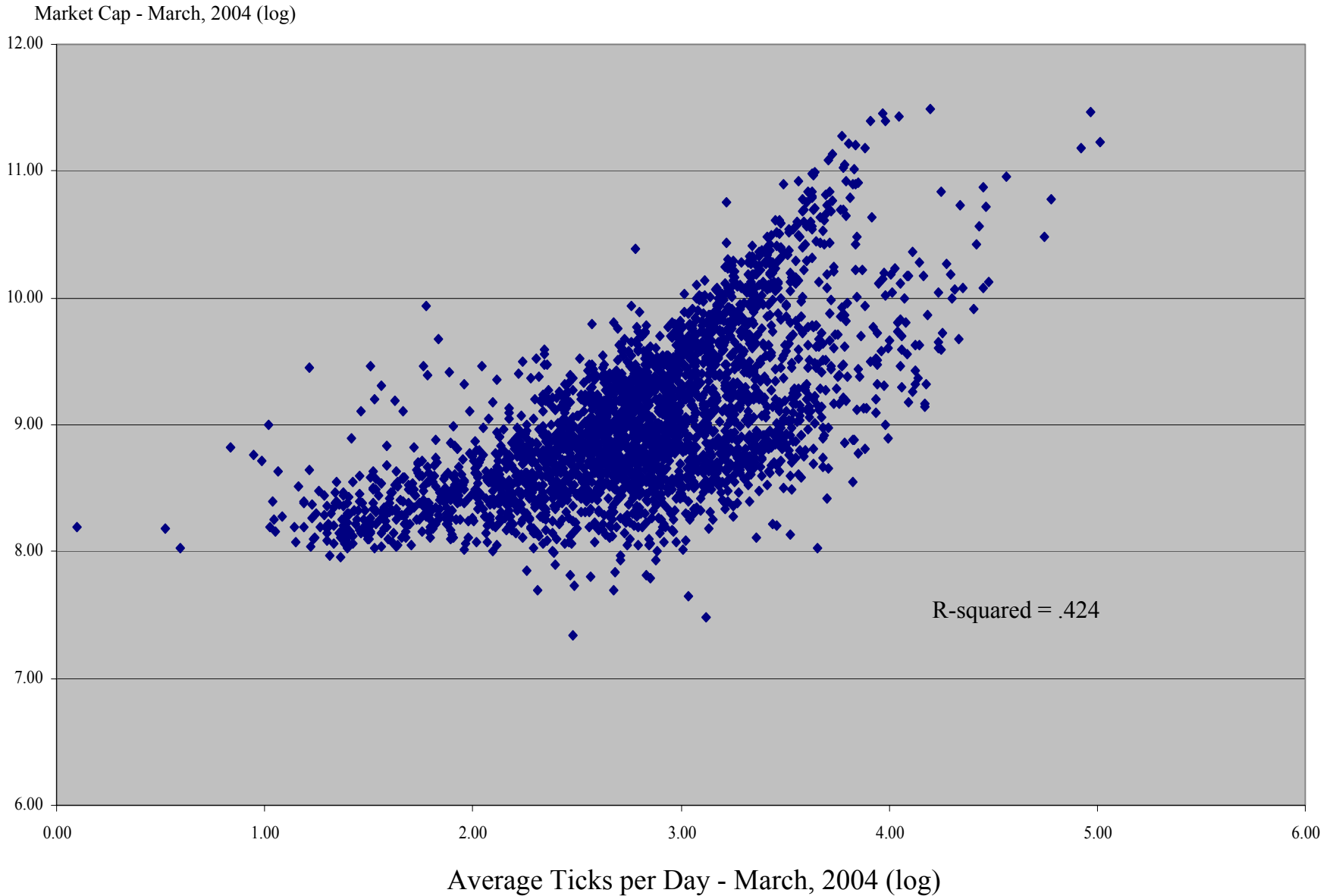
Challenges

- The asynchronous nature of tick data.
- Differences in tick frequency across a universe of securities.
- Intraday patterns in tick frequency.
- The myriad of possible error types.
- Multiple sources of information (floor versus regional exchanges).
- Beginning of day effect; priming filters and corporate actions.
- Speed.

Relationship between Ticks and Volume Russell 3000 Constituents March, 2004

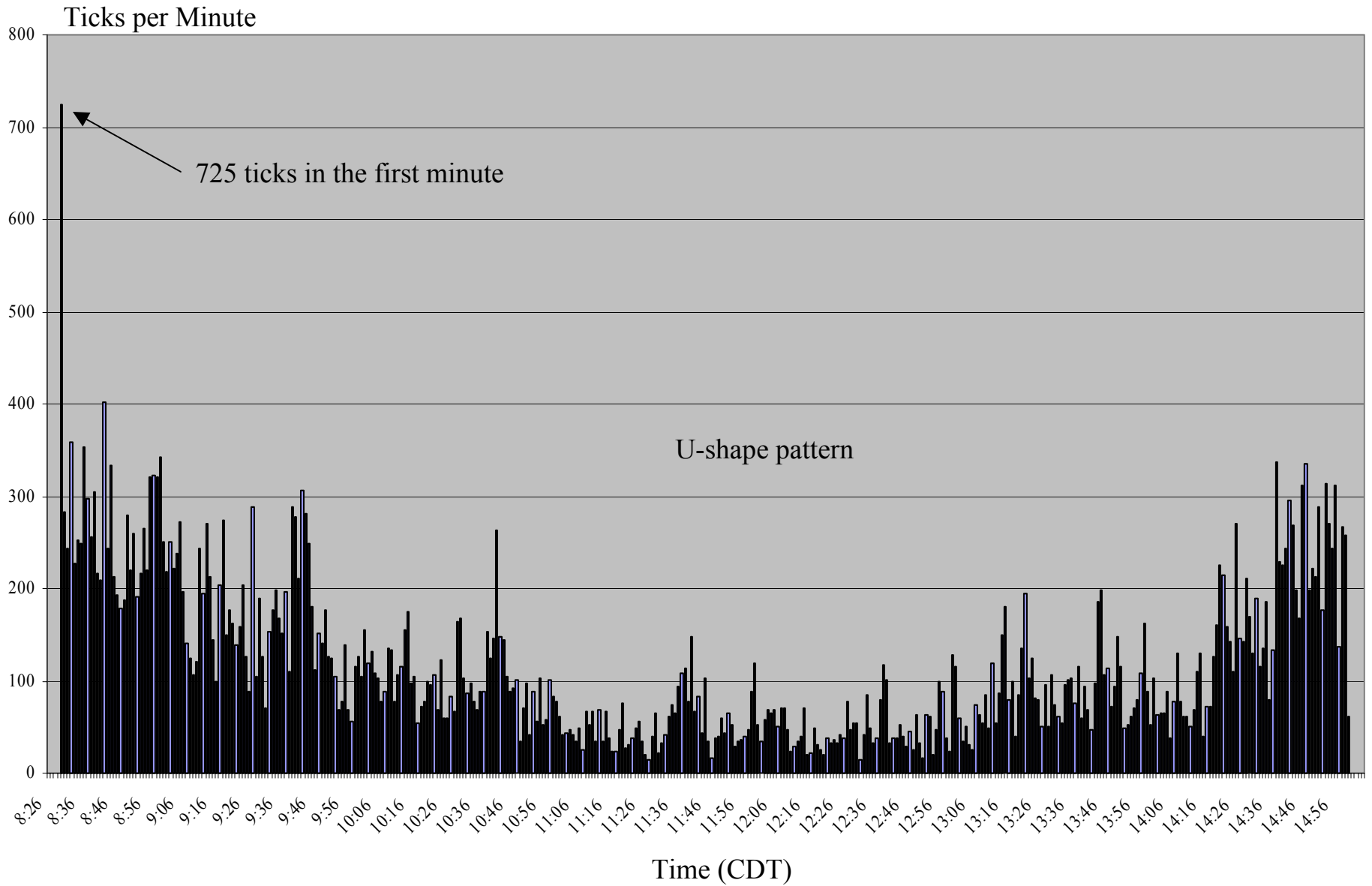


Relationship between Ticks and Market Capitalization
Russell 3000 Constituents
March, 2004



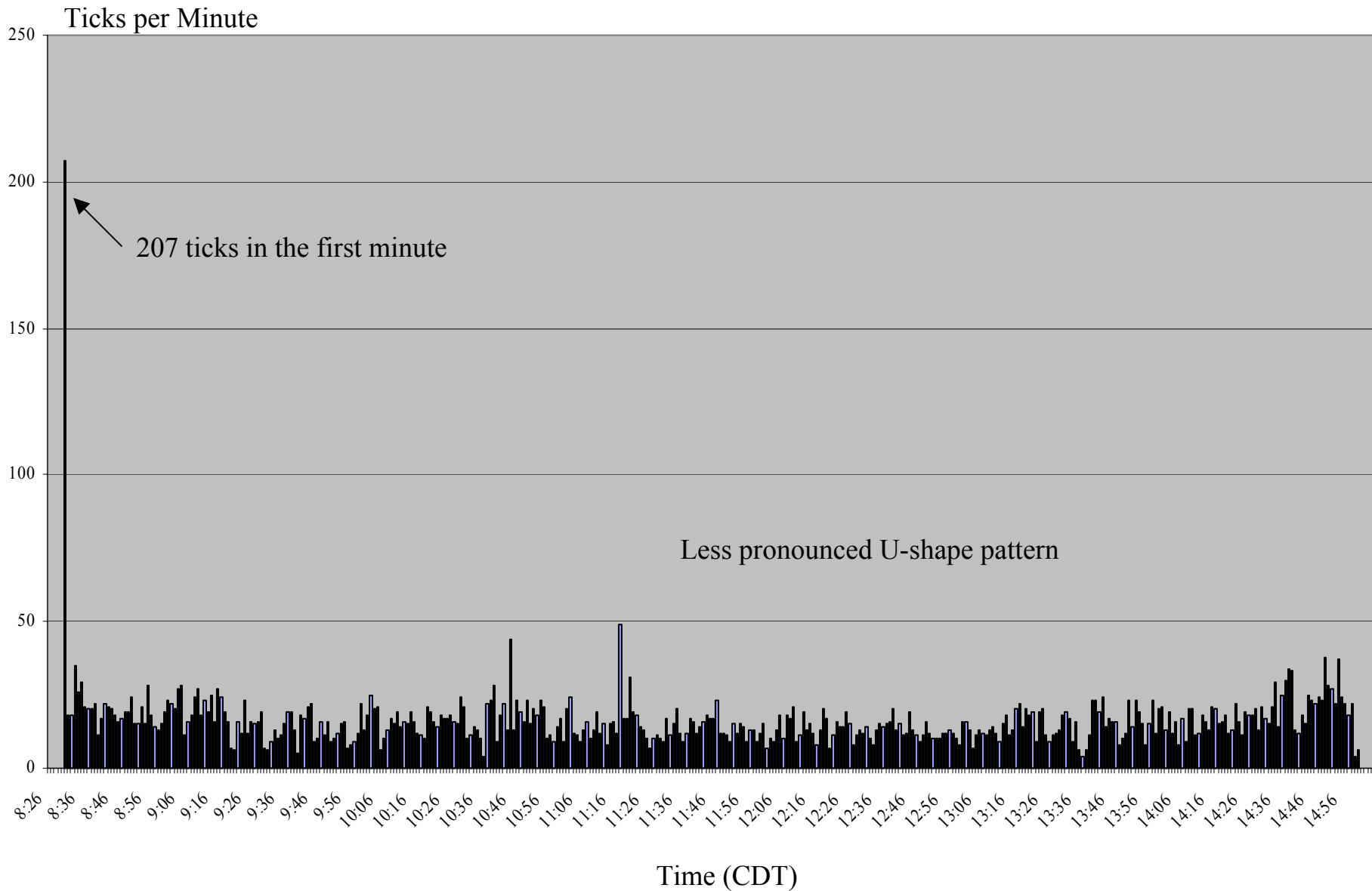
Group I

High Volume NASDAQ Issues (MSFT, INTC, CSCO)



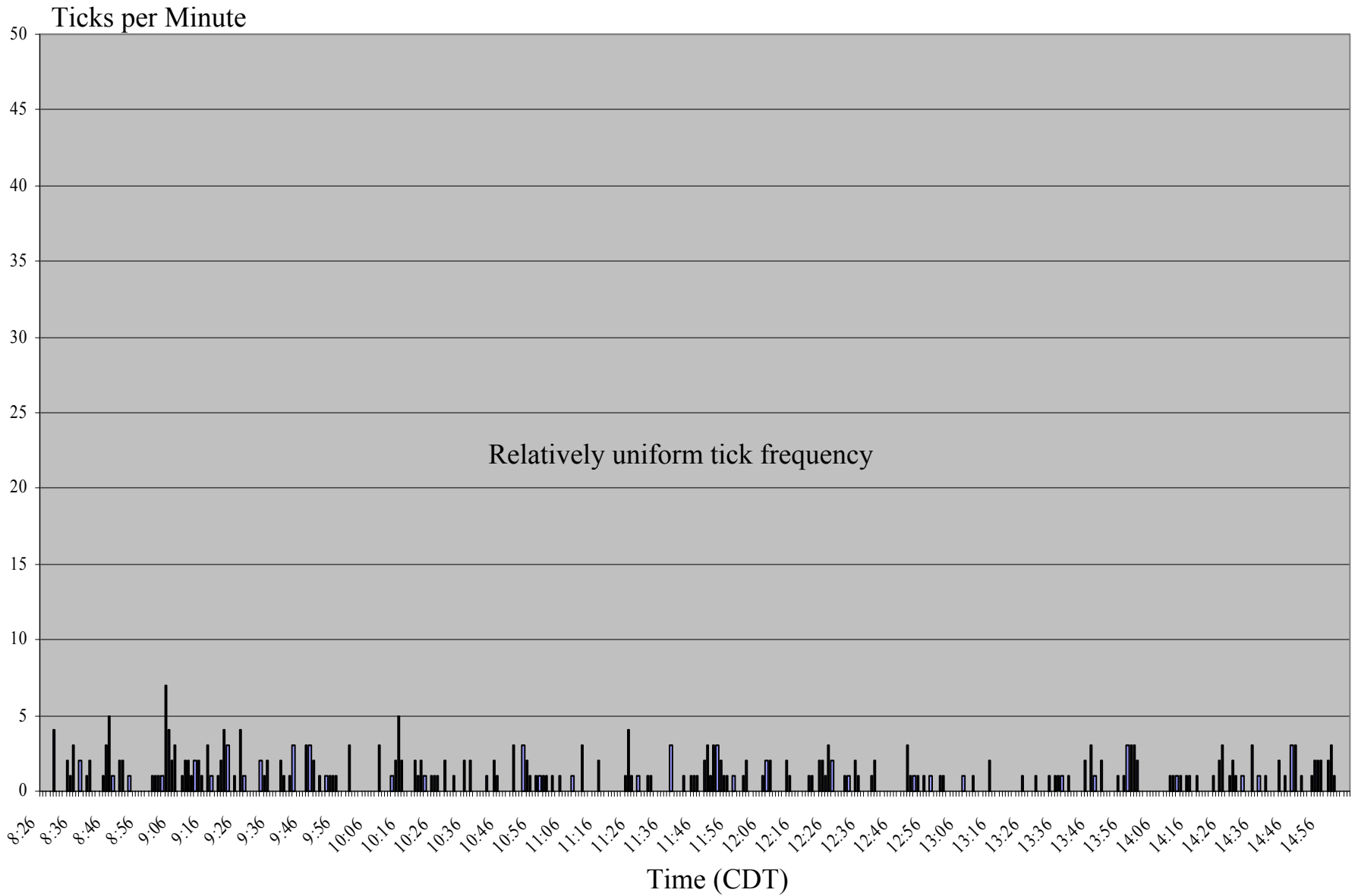
Group II

Large Cap NYSE Issues (GE, PFE, JNJ)



Group III

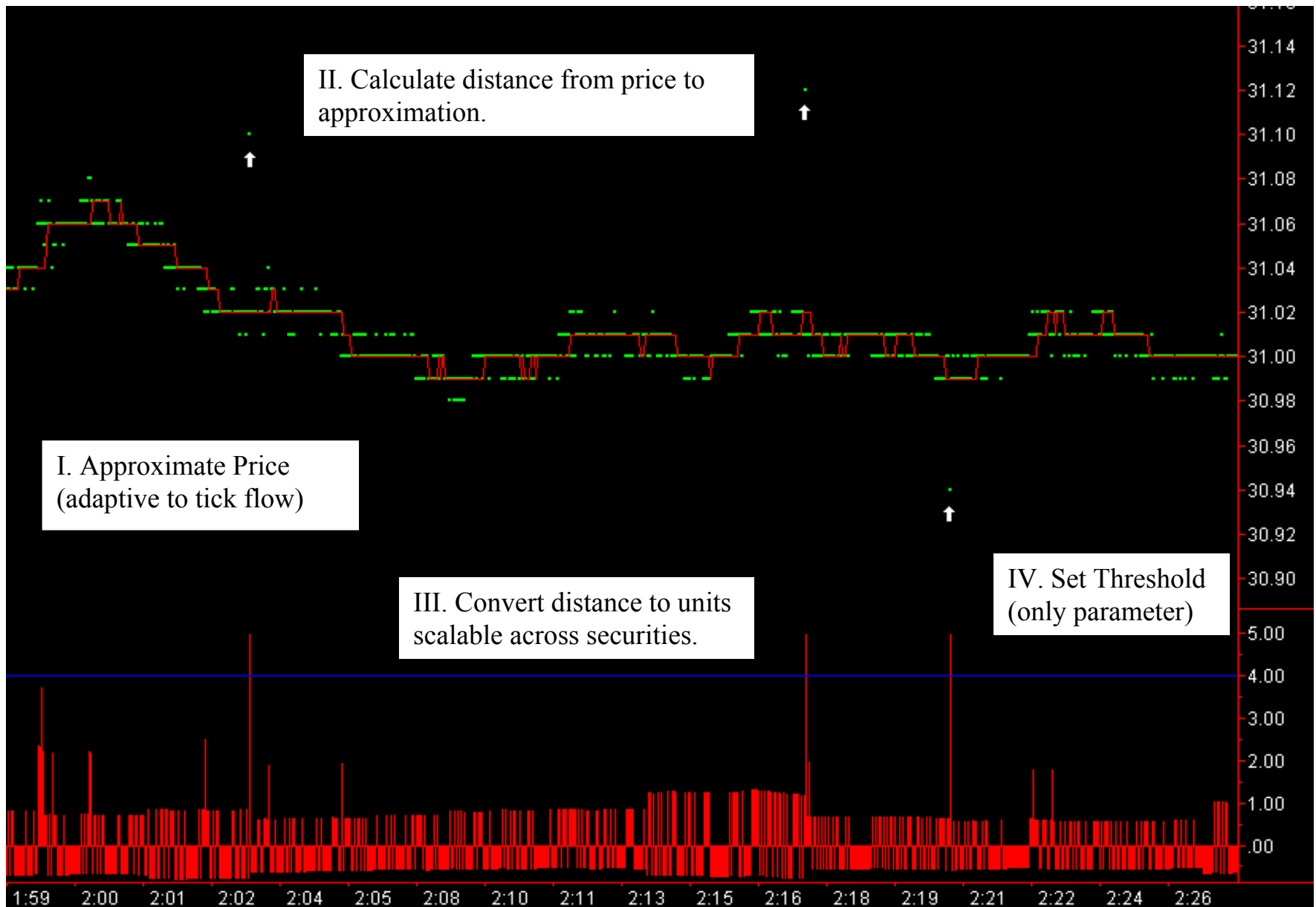
Small and Mid Cap Issues (IOM, RESP)



Solutions

- Tick filters must be highly adaptive to tick flow:
 - across securities, and
 - throughout the day within the same security.
- Tick filters must not introduce excessive delay from computation time or the need for confirming data points.
- Tick filter should not treat historical data different from real-time data.
- Tick filter should have minimal parameters.

The Tick Data, Inc. Filter



Market Data

- Trades, Quotes
- Level2 Quotes, Orders
- 100,000 records per second
- 1,000,000 attributes per second
- 100,000,000 calculations per second?

Continuous Queries

Proprietary ..

- Moving averages
- Moving vwaps
- Indices
- Nbbo
- Size at inside
- Weighted sizes at bid and ask

Sample Clients

Given table

```
x: ([]time;sym;price;size)
```

- All: t,:x
- Last: t,:select last price,
last size by sym from x
- vwap: t+:select size wsum
price, sum size by sym from x

Corporate Actions

- 5 terabytes of history
- 100's of daily indicators