

# Depth and Volume

August 27, 2015

## 1 Data

We use historical data from the NYSE Trade and Quote database<sup>1</sup>, made available by Wharton Research Data Services. The dataset “contains intraday transactions data (trades and quotes) for all securities listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), as well as Nasdaq National Market System (NMS) and SmallCap issues”<sup>3</sup>.

Our data cover all trades for a sample of symbols<sup>4</sup> over the period of Jan 1, 2014 - Dec 31, 2014. No time filtering is used, and we are left with a total of 249 trading days.

### 1.1 Trades Data

The trading data is organized as a sequential collection of trades, where a trade is defined as a transaction with a unique trade sequence number. A trade must be between no more than two parties, in one symbol, at one price, on one exchange, at one time.

Since we filter out some special cases (trade corrections, duplicates, etc.), we will not go into that here.<sup>5</sup>

---

<sup>1</sup>We note that this data is known to have inaccuracies regarding quotes’ timestamps, and such issues may carry over to trade timestamps. These issues stem from timestamps originating from a secondary source (i.e., not the exchanges’ order matching engines) and being derived after an exchange aggregation method which introduces further latency<sup>2</sup>.

<sup>3</sup>See <https://wrds-web.wharton.upenn.edu/wrds/ds/taq/index.cfm>.

<sup>4</sup>Our sample of symbols consists of AMD, BAC, C, GOOG, GRPN, JBLU, MSFT, and RAD. We would like to include BRKA and BRKB, but millisecond TAQ data does not exist for them. Strange.

<sup>5</sup>[www.nyxdata.com/doc/224904](http://www.nyxdata.com/doc/224904)

### 1.1.1 Trades Against One Counterparty

Alice buys 50 shares of GOOG for \$700 on NASDAQ at 11:00 against a corresponding quote of at least 50 shares. The trades data would show the following:

Time	Symbol	Volume	Price	Exchange
11:00:00	GOOG	50	\$700.00	NASDAQ

### 1.1.2 Trades Against Multiple Counterparties

Alice sells 50 shares of GOOG for \$700 on NASDAQ at 11:00, but the first-in-line quote is only big enough for 10 shares. The next two quotes, however, are of size 6 and 80, respectively, and so Alice can be accommodated by these three quotes. The trades data would show the following:

Time	Symbol	Volume	Price	Exchange
11:00:00	GOOG	15	\$700.00	NASDAQ
11:00:00	GOOG	6	\$700.00	NASDAQ
11:00:00	GOOG	29	\$700.00	NASDAQ

### 1.1.3 Trades On Multiple Exchanges

Alice sends a buy order 20 shares of GOOG for \$700 on NASDAQ and a buy order 30 shares of GOOG for \$700 on NYSE. Furthermore, the first-in-line quote on NYSE is only big enough for 1 share, but the next quote in the NYSE queue is for 100 shares. The trades data would show the following:

Time	Symbol	Volume	Price	Exchange
11:00:00	GOOG	20	\$700.00	NASDAQ
11:00:00	GOOG	1	\$700.00	NYSE
11:00:00	GOOG	29	\$700.00	NYSE

### 1.1.4 Trades At Multiple Prices

Alice sends a market buy order for 50 shares of GOOG on BATS BYX. The first-in-line quote is to sell 3 shares at \$700, and the next quote in the queue is for 16 shares at \$700.01. The next quote is for 19 shares at \$700.01, and the following quote is for 60 shares at \$700.03. The trades data would show the following:

Time	Symbol	Volume	Price	Exchange
11:00:00	GOOG	3	\$700.00	BATS BYX
11:00:00	GOOG	16	\$700.01	BATS BYX
11:00:00	GOOG	19	\$700.01	BATS BYX
11:00:00	GOOG	12	\$700.03	BATS BYX

### 1.1.5 Order Gets Routed

Alice sends a market sell order for 50 shares of GOOG on BATS BYX. The first-in-line quote is to buy 5 shares at \$700, and the next quote in the queue is for 60 shares at \$699.98. However, NASDAQ is the national best bid with a size of 1 share at \$699.99. By Reg NMS, Alice's sell order gets routed to NASDAQ and she sells 1 share for \$699.99. Now, the national best bid is the 60 shares at \$699.98 on BATS BYX. Alice's order transacts her remaining 44 shares against the 60 share bid on BATS BYX. The trades data would show the following:

Time	Symbol	Volume	Price	Exchange
11:00:00	GOOG	5	\$700.00	BATS BYX
11:00:00	GOOG	1	\$699.99	NASDAQ
11:00:00	GOOG	44	\$699.98	BATS BYX

### 1.1.6 Trades At Different Times

Alice buys 40 shares of GOOG for \$700 on BATS BYX at 11:00:00 against two orders of size 13 and 90, respectively. Bob then sells 10 shares of GOOG for \$699.97 on NASDAQ at 11:00:04. The trades data would show the following:

Time	Symbol	Volume	Price	Exchange
11:00:00	GOOG	13	\$700.00	BATS BYX
11:00:00	GOOG	27	\$700.00	BATS BYX
11:00:04	GOOG	10	\$699.99	NASDAQ

## 1.2 Quotes Data

The quotes data is organized as a sequential collection of quote updates, which only occur when the best bid/offer (BBO) has changed. These changes could be the result of a trade, a quote addition, a quote retraction, or a quote modification.

The following scenarios are all for one symbol on one exchange, as their quotes data are independent of one another.

### 1.2.1 Nothing Happens

The market opens at 09:30 with a \$9.99 bid for 30 shares and a \$10.03 bid for 15 shares. This remains the BBO for the remainder of the day:

Time	Bid Size	Bid	Offer	Offer Size
09:30:00	30	\$9.99	\$10.03	15

### 1.2.2 Improved Quotes

The market opens at 09:30 with a \$9.99 bid for 30 shares and a \$10.03 bid for 15 shares. At 09:44:00, a new \$10.00 bid for 2 shares enters the market. At

10:20:00, a new \$10.02 offer for 30 shares enters the market. This remains the BBO for the remainder of the day:

Time	Bid Size	Bid	Offer	Offer Size
09:30:00	30	\$9.99	\$10.03	15
09:44:00	2	\$10.00	\$10.03	15
10:20:00	2	\$10.00	\$10.02	30

### 1.2.3 Joined Quotes

The market opens at 09:30 with a \$9.99 bid for 30 shares and a \$10.03 bid for 15 shares. At 09:48:00, a new \$9.99 bid for 20 shares enters the market, joining the existing bid for 30 shares. This remains the BBO for the remainder of the day:

Time	Bid Size	Bid	Offer	Offer Size
09:30:00	30	\$9.99	\$10.03	15
09:48:00	50	\$9.99	\$10.03	15

### 1.2.4 Retracted (or Traded) Quotes

The market opens at 09:30 with two \$9.99 bids for 10 and 20 shares, respectively, and a \$10.03 bid for 15 shares. At 09:33:00, the \$9.99 bid for 10 shares is retracted (or traded). At 09:34:00, the \$9.99 bid for 20 shares is retracted (or traded), and now the previously existing \$9.98 bid for 100 shares is the new best bid. This remains the BBO for the remainder of the day:

Time	Bid Size	Bid	Offer	Offer Size
09:30:00	30	\$9.99	\$10.03	15
09:33:00	20	\$9.99	\$10.03	15
09:34:00	100	\$9.98	\$10.03	15

## 2 Transforming Data

Each data point concerned a specific symbol  $i$ 's activity on a specific exchange  $j$  for a specific date  $t$ .

### 2.1 Volume

Volume  $Q_{ijt}$  is computed as the sum of the volumes  $q$  of trades occurring between 10:00 and 15:30 in symbol  $i$  on exchange  $j$  on date  $t$ :

$$Q_{ijt} = \sum_{q \in ijt} q$$

## 2.2 Depth

We begin by computing the depth of each quote update, indexed by  $u$ . We used two different methods.

### 2.2.1 Average Method

This method measures the quote update’s depth  $\overline{q}_u$  as the average of the size of the best bid  $q_{bu}$  and the size of the best ask  $q_{au}$ :

$$\overline{q}_u = \frac{q_{bu} + q_{au}}{2}.$$

### 2.2.2 Min Method

This method measures the quote update’s depth  $\overline{q}_u$  as the minimum between the size of the best bid size  $q_{bu}$  and the size of the best ask  $q_{au}$ :

$$\overline{q}_u = \min(q_{bu}, q_{au}).$$

### 2.2.3 Combining All $\overline{q}_u$ Into One Measure

For each quote update  $u$ , we compute its duration  $\Delta T_u$ —that is, how long this quote update  $u$  represents the state of the book. We do this by subtracting the timestamp  $T_u$  of quote update  $u$  from the timestamp  $T_{u+1}$  of the next quote update  $u + 1$ :

$$\Delta T_u = T_{u+1} - T_u.$$

To summarize the depth  $\overline{d}_{ijt}$  between 10:00 and 15:30 in symbol  $i$  on exchange  $j$  on date  $t$ , we take the sum of  $\overline{q}_u$  weighted by their corresponding  $\Delta T_u$ .

$$\overline{d}_{ijt} = \sum_{u \in ijt} q_u \cdot \Delta T_u$$

John mentioned how NASDAQ quotes effectively treat its different DMMs as different exchanges; make sure to do special logic for NASDAQ.

## 3 Results

Due to wide heterogeneity in volumes between symbols, we take the logarithm of both depth and volume.

We filter out data points with 0 depth, as this implies empty bids or asks. (Penalizing quotes past the NBBO should solve this problem.)

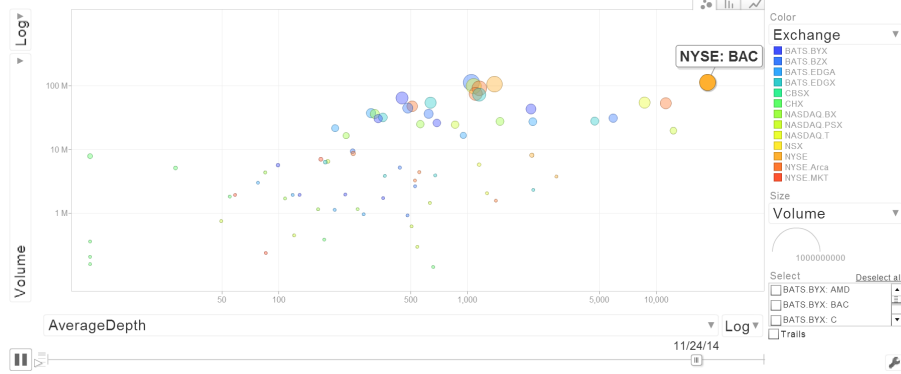


Figure 1: The values of daily depth and volume are plotted against each other across symbols. The colors correspond to different exchanges.

Graphs are located in Dropbox<sup>6</sup>.

From plotting the different symbols together, we observed that the depth v. volume relationship appeared to be segmented by symbol. Thus, the following analyses separates symbol.

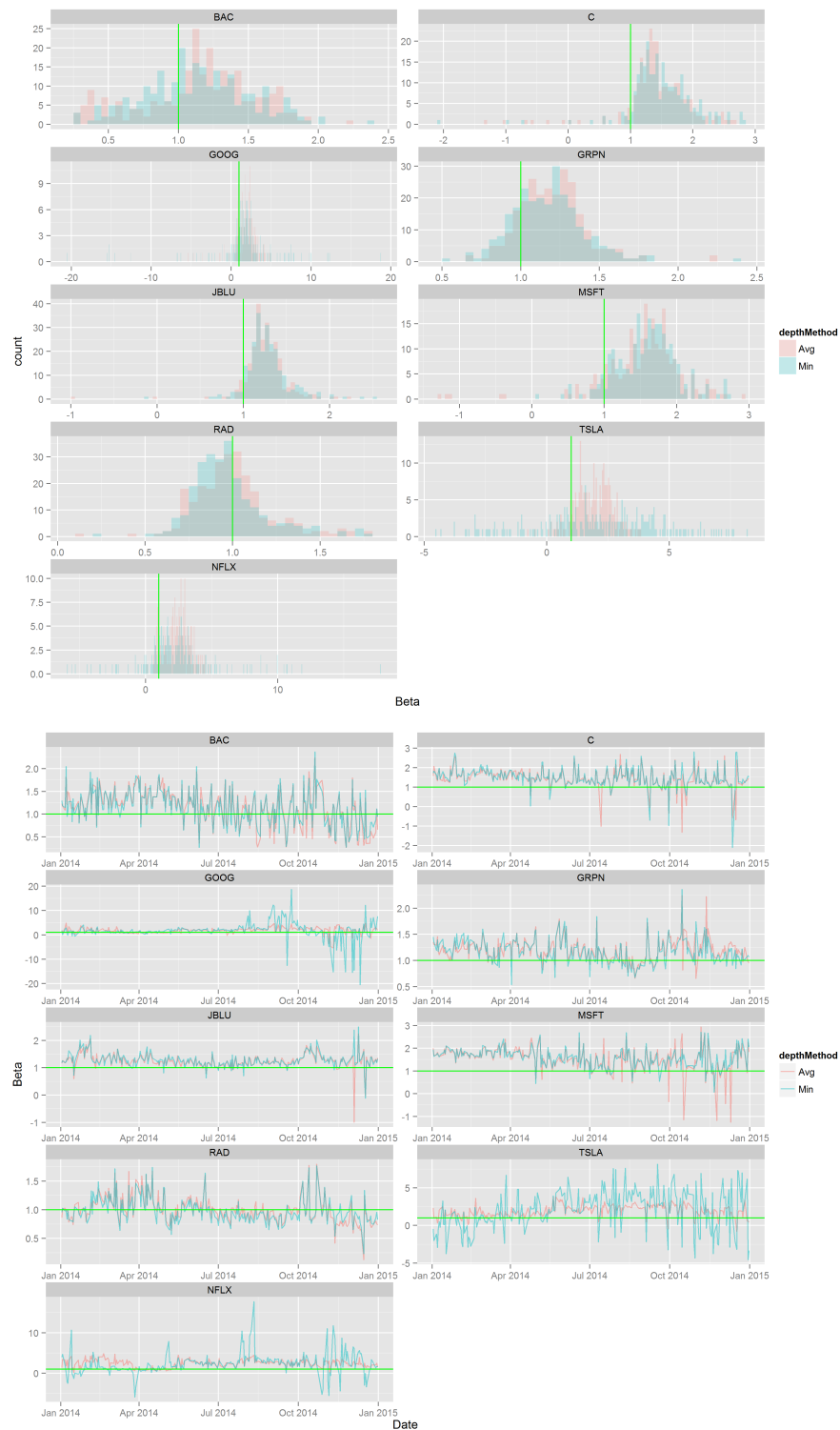
For each date  $t$  and symbol  $i$ , we fit the regression model:

$$\log \bar{d}_{ijt} = \alpha_{it} + \beta_{it} \cdot \log Q_{ijt} + \epsilon_{ijt}.$$

We ran the above regressions for all 4 combinations of:

- Min or Average method of measuring depth
- Filtering or Unfiltering the four small exchanges (CBSX, CHX, NSX, NASDAQ PSX), as they account for less than %5 of total volume.

<sup>6</sup><https://www.dropbox.com/sh/kxtlz8mb4lzfvgl/AAACcauEobiX21IBCM...6IUja?dl=0>



7

Figure 2: Unfiltered dataset.

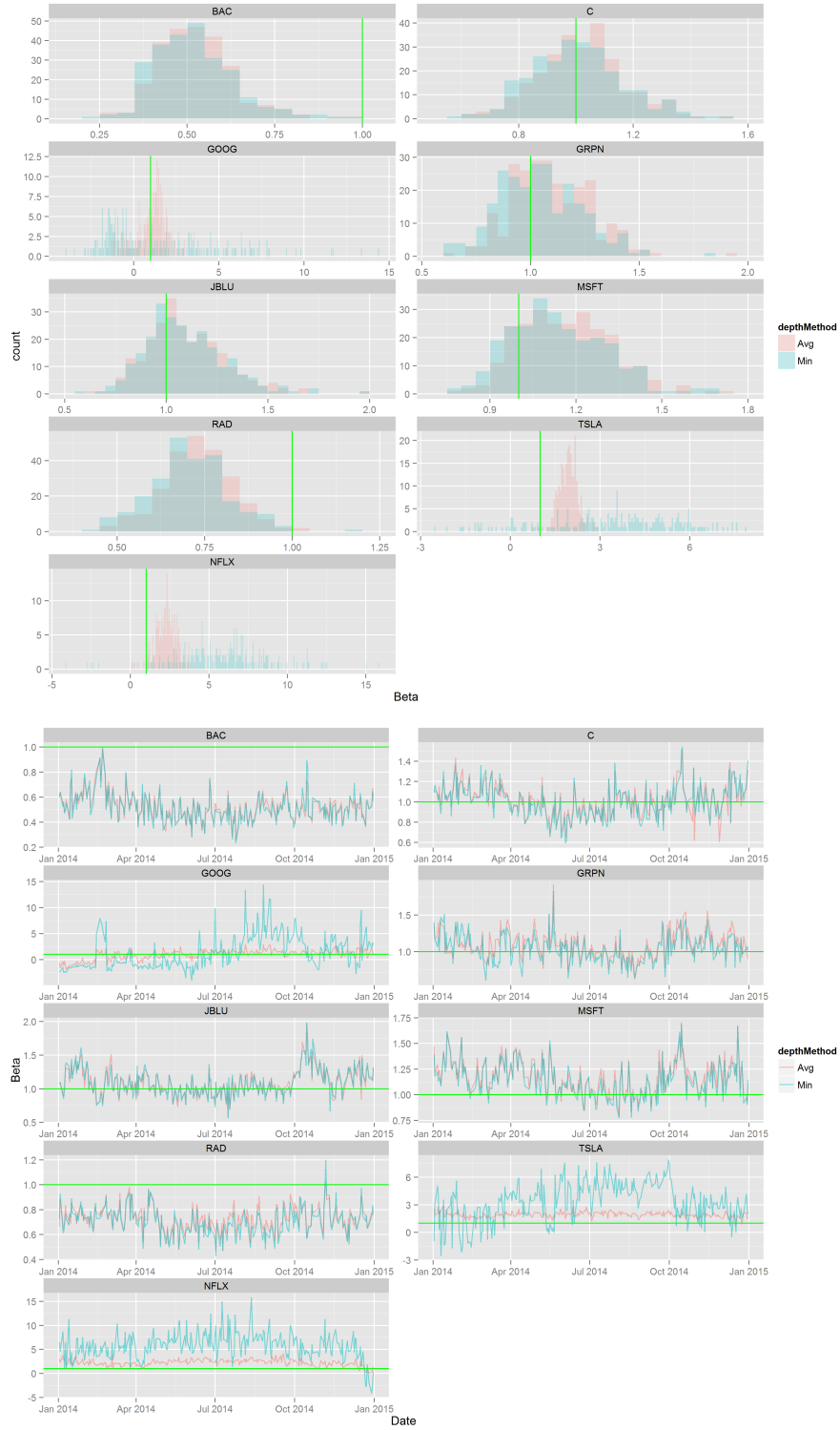


Figure 3: Filtered dataset (excluded CBSX, CHX, NSX, NASDAQ PSX).



## 3.1 Curious Behavior

### 3.1.1 Low Depth and High Volume in CHX

We investigate the CHX points which occur in the top right region (low depth, high volume).

For one such outlier, BAC on CHX on 10-Jan-2014, we open the individual trades: [[TRADE DATA OF ALL CHX TRADES IN BAC ON JAN 10]]

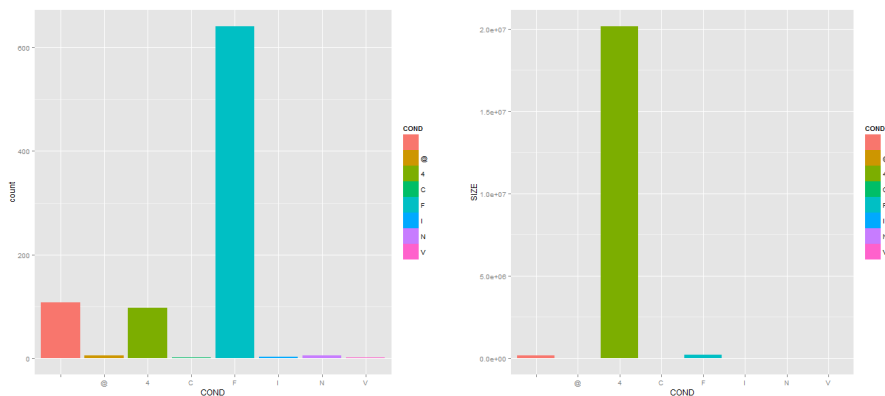


Figure 4: In January, although trades with condition code 4 constituted less than 15% of the trades, over 95% of volume held this condition code.

Most interestingly, we observe two large trades with condition code 4, indicating "Derivatively Priced" trades. Online literature suggests that activity in Exchange Traded Products<sup>7</sup> largely make up these trades, perhaps when different symbols are negotiated at a package price and the individual trades semi-arbitrarily take on parts of the cost.

The fact that such trades are not considered to be a true "Last Sale"<sup>8</sup> and are allowed outside of limit-up and limit-down prices<sup>9</sup> strongly indicates that these trades' prices are not dictated by the BBO.

As a response, all trades with condition code 4 were filtered out.

### 3.1.2 Flat Relationship of Volume v. Depth for BAC

This flat relationship could be attributable to the CHX outlier which hovered in the low-depth/high-volume region. Filtering out Derivatively Priced trades remedied this flatness.

<sup>7</sup><http://www.investopedia.com/terms/e/exchange-traded-products-etp.asp>

<sup>8</sup><http://www.nasdaqtrader.com/content/technicalsupport/specifications/utp/utdfspection.pdf>

<sup>9</sup><http://www.finra.org/industry/trf/trade-report-modifiers-and-applicability-limit-uplimit-down-luld-price-bands>

### 3.1.3 GOOG: Different Volumes for Similar Depths

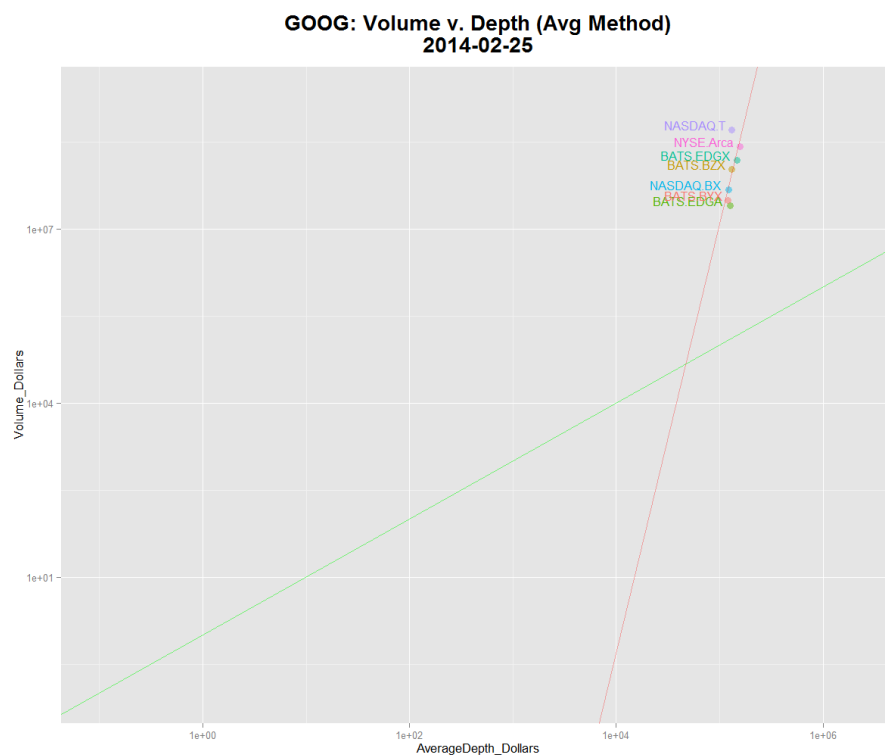


Figure 5: A sample date in GOOG exhibiting both low variability in depth, and BATS EDGA and BATS BYX pushing the regression line to a negative slope.

First we address the low variability in depth.

Normally, a symbol will have a quote history like where spreads are mostly one tick wide, and exchanges have similar BBOs. Because of this, the differences between exchanges' depths are largely dictated by the size shown on the bid and offer.

DATETIME	SYMBOL	EX	BIDSIZ	BID	OFR	OFRSIZ
2014-01-27 10:44:37	JBLU	T	43	8.64	8.65	42
2014-01-27 10:44:37	JBLU	B	11	8.64	8.65	12
2014-01-27 10:44:37	JBLU	P	11	8.64	8.65	23
2014-01-27 10:44:37	JBLU	Y	3	8.64	8.65	5
2014-01-27 10:44:37	JBLU	Y	3	8.64	8.65	4
2014-01-27 10:44:37	JBLU	A	1	8.64	8.66	1
2014-01-27 10:44:37	JBLU	A	1	8.64	8.66	1
2014-01-27 10:44:37	JBLU	X	2	8.63	8.65	1
2014-01-27 10:44:37	JBLU	Y	3	8.64	8.65	3
2014-01-27 10:44:37	JBLU	P	11	8.64	8.65	18
2014-01-27 10:44:37	JBLU	C	5	8.63	8.65	2
2014-01-27 10:44:37	JBLU	C	5	8.63	8.65	1

On the other hand, GOOG's quote behavior is comprised of multiple tick-wide spreads, and there is much heterogeneity in the exchanges' BBOs. Because of this, the differences between exchanges' depths are not sufficiently captured by our current measure, since our measure does not take into account spread.

DATETIME	SYMBOL	EX	BIDSIZ	BID	OFR	OFRSIZ
2014-01-27 10:00:16	GOOG	X	1	1109.12	1114.25	1
2014-01-27 10:00:16	GOOG	Y	1	1110.9	1112.47	1
2014-01-27 10:00:17	GOOG	T	2	1110.9	1113.35	1
2014-01-27 10:00:17	GOOG	P	4	1110.21	1113.26	4
2014-01-27 10:00:17	GOOG	Y	1	1110.9	1113.27	6
2014-01-27 10:00:17	GOOG	Y	1	1110.9	1113.27	3
2014-01-27 10:00:17	GOOG	Y	1	1110.9	1159.74	1
2014-01-27 10:00:17	GOOG	K	1	1110.9	1116.03	1
2014-01-27 10:00:17	GOOG	J	3	1110.4	1159.74	1

To further support this, we look include two more high-priced stocks, NFLX and TSLA. We observe wide, inconsistent spreads among their different exchanges' BBOs. Indeed, we see the same behavior in their slopes.

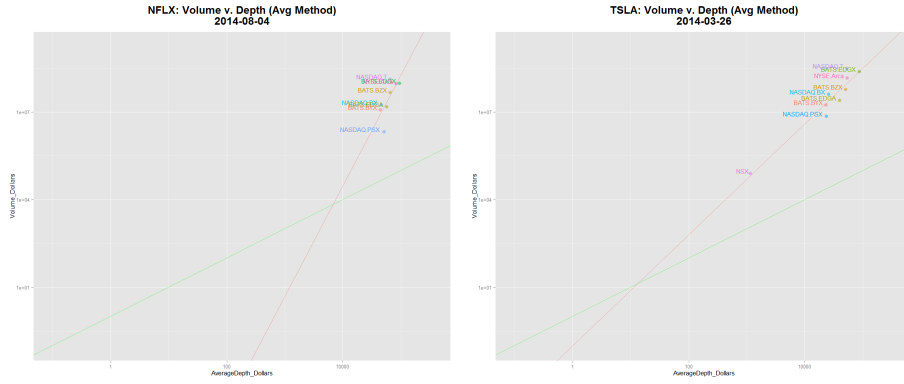


Figure 6: A sample date in GOOG exhibiting both low variability in depth, and BATS EDGA and BATS BYX pushing the regression line to a negative slope.

Finally, we look at BYX's skewing of the slope, and we note interesting clumping of higher BBO sizes in BYX:

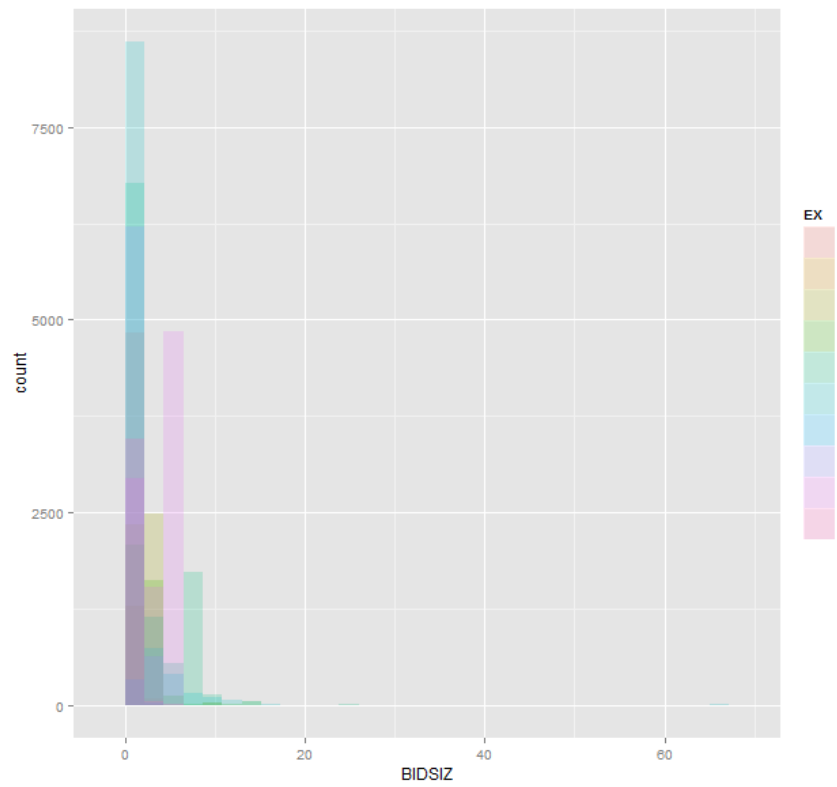


Figure 7: Histogram of bid size, note the bump in BYX being at 7. Why?

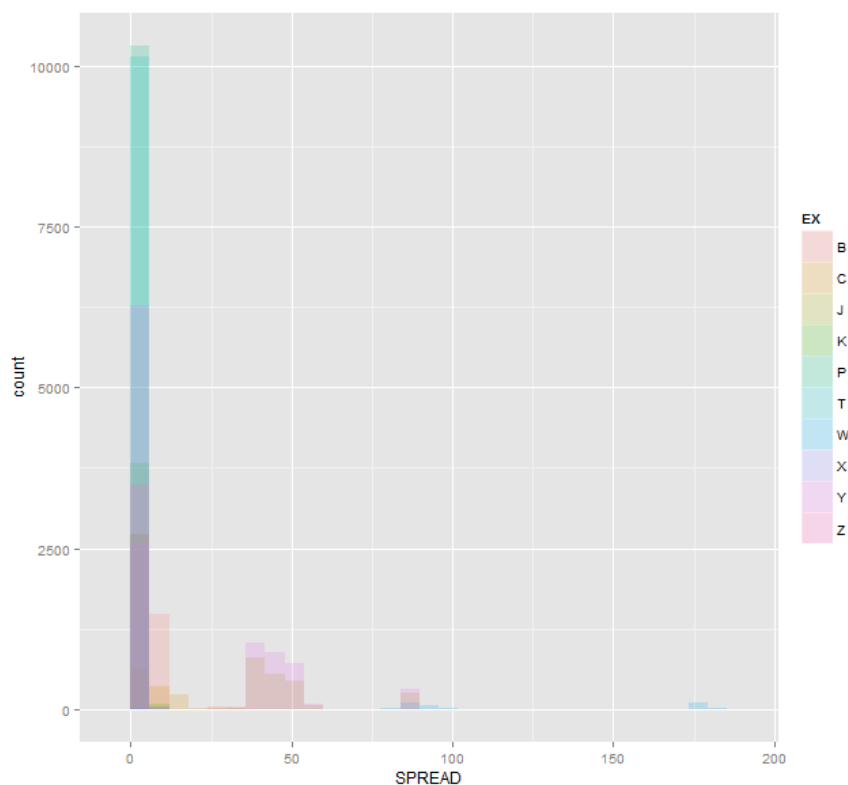


Figure 8: Histogram of spread, note large spread in BYX around 40. BYX has higher spreads. See if bump only happens on certain days or is always hinted at by data.

### 3.1.4 Depth Inconsistency and its Consequences

In a stock like BAC, the book is typically one tick wide with volume on the order of dozens:

DATETIME	SYMBOL	EX	BIDSIZ	BID	OFR	OFRSIZ
2014-03-20 10:02:00	BAC	T	64	17.51	17.52	257
2014-03-20 10:02:00	BAC	Z	19	17.51	17.52	103
2014-03-20 10:02:00	BAC	P	73	17.51	17.52	186
2014-03-20 10:02:00	BAC	K	33	17.51	17.52	218
2014-03-20 10:02:00	BAC	N	6	17.51	17.52	437

However, exchanges like CBSX, CHX, NSX, and NASDAQ PSX generally have wider spreads and significantly less size on their BBOs. This leads to their distance from the other exchanges in depth/volume space, and they exert a disproportionate amount of influence on the regression.

Consider the following quotes from NSX

DATE TIME	SYMBOL	EX	BIDSIZ	BID	OFR	OFRSIZ
2014-03-20 10:02:07	BAC	C	1	17.5	17.52	2
2014-03-20 10:02:07	BAC	C	1	17.5	17.52	1
2014-03-20 10:02:07	BAC	C	1	17.5	0.00	0
2014-03-20 10:02:07	BAC	C	1	17.5	17.56	1
2014-03-20 10:02:08	BAC	C	1	17.5	17.51	1
2014-03-20 10:02:08	BAC	C	1	17.5	17.56	1

This sort of quote behavior is not present day-to-day, but gave us reason to run a separate version of our analyses without the four exchanges.