

IEX Market Microstructure Analysis

GROUP MEMBERS:

SITENG JIN

HENG LU

YAQI WU

JIEYUN WANG

MIAOSHI LI

INSTRUCTORS:

PACA NGUYEN

Ph. D. MORTON N. LANE

December 1, 2014

Contents

1. Introduction.....	1
1.1. Background	1
1.2. Research purpose.....	1
1.3. Data Scope.....	1
1.4. Hardware and Software.....	2
2. Data.....	3
2.1. Data Collection	3
2.2. Data Processing.....	4
2.3. Computational Improvement	5
3. Statistical Analysis.....	6
3.1. Assumptions.....	6
3.2. Two Sided T-Test.....	6
3.3. Regression Models	7
3.4. Presentation of Models.....	8
4. Summary.....	10
5. Reference.....	12

1. Introduction

1.1. Background

Back on October 25, 2013, a small dark pool named “IEX Group Inc.”, which is known as an alternative trading system (ATS) was established in New York City. Since then, IEX has seen steady growth in trading volume and it has been making effort to convert itself to a public stock exchange.

The major difference between dark pool and stock exchange is that dark pool is a private trading venue where buy and sell orders are not displayed to the public. Thus, exchange is more heavily regulated and more costly to operate. Compared to existing stock exchanges, IEX allows for a lower trading fees; it gives broke-dealer priority to jump to the front of the trading queue, so that their orders would get a greater chance to be filled. According to Tabb Group, “ About 40% of stock trading in the U.S is traded in the dark pool” (Bradley Hope, IEX Pricing Aims to Drain ‘Dark Pools’, wsj.com). In seeking exchange status, IEX is looking for new revenue sources.

1.2. Research purpose

The main purpose of our practicum is desired to evaluate the market characteristics of a sample set of stocks nine months before (January 2013 – October 2013) and nine months after (October 2013 – August 2014) the opening of IEX. However, because of the running speed of our machine and time limit, we finally used data of two month before and two month after IEX’s presence to do analysis.

Besides, through statistic analysis, we plan to evaluate changes in various market factors, including liquidity, trading volume, price, and so on, of the sample set of stocks after they were traded on IEX.

1.3. Data Scope

To detect that whether IEX has an influence on the microstructure of the market, we use NYSE Trade and Quote (TAQ) Data of four months for the analysis. More specifically, it contains Daily Trade files and Daily NBBO Files from September 9, 2013 to December 6, which is about 2 month before IEX establishment and 2 month after IEX establishment

(IEX started on Oct 25, 2013). The NYSE TAQ Data contains the trading information on all 19 exchanges listed in Table 1.3.1. We focused on the performance of the most popular stocks – top 10 stocks from S&P 500, due to the market capitalization, shown in Table 1.3.2.

List of Exchanges
NYSE MKT Stock Exchange
NASDAQ OMX BX Stock Exchange
National Stock Exchange
FINRA
International Securities Exchange
Direct Edge A Stock Exchange
Direct Edge X Stock Exchange
Chicago Stock Exchange
New York Stock Exchange
NASDAQ OMX Stock Exchange
NYSE Arca SM
Consolidated Tape System
NASDAQ Stock Exchange
CBOE Stock Exchange
NASDAQ OMX PSX Stock Exchange
BATS Y-Exchange
BATS Exchange

Table 1.3.1 List of Exchanges

Symbol	Name
AAPL	Apple Inc.
XOM	Exxon Mobil Corp.
GOOG	Google Inc.
MSFT	Microsoft Corp.
JNJ	Johnson & Johnson
WFC	Wells Fargo
GE	General Electric
WMT	Wal-Mart Stores
CVX	Chevron Corp.
JPM	JPMorgan Chase & Co.

Table 1.3.2 Stocks Selected

1.4. Hardware and Software

Since the trade and NBBO data for one day is about 20 gigabytes, to process all data for the whole four month requires a computer with significant processing power and a substantial memory. As a result, we made use of the server, purchased by BOX, in the MSFE Department office. The server ran on a 3.2-gigahertz microprocessor and had approximately 4 terabytes of storage capacity.

The huge amount of data was loaded into a Microsoft SQL Server database and then

filtered and aggregated by various attributes for further analysis in the database management tool, SQL Server Management Studio. Furthermore, a SQL Server Integration Services (SSIS) package was created to manage the data processing automatically in SQL Server Data Tools for Visual Studio 2012. For more advanced statistical analysis concerning the influence of IEX, such as individual t-test, linear regression and time series analysis, we used a statistical software -- R studio and Excel.

2. Data

2.1. Data Collection

In order to dig into the changes in market characteristics after the opening of IEX, we need to discover if there is any change in the market microstructure for different exchanges and different stocks from a statistical point of view. And try to identify if those changes are caused by IEX or other macro market factors.

The daily Trade file contains every trade reported to the consolidated tape from all CTA participants. Every trade recorded identifies the time, exchange, security, volume, price, sale condition, and more. The daily NBBO file contains continuous National Best Bid and Offer updates and consolidated trades and quotes for all listed and non-listed issues. The raw data was constructed with rows of character strings for different recorded time. We followed the Daily TAQ Client Specification of the dataset and extracted certain properties that we need for data processing. The specifics of data content extracted from daily Trade and NBBO file for the project is shown in Table 2.1 and Table 2.2.

Field Name	Data Type	Description
Time	Date	Time of trade (HHMMSSXXX)
Exchange	Text	The Exchange where trade took place
Symbol	Text	Stock symbol
Sale Condition	Text	Type of sale condition
Trade Volume	Number	The total volume of shares traded in the trade
Trade Price	Number	The price of trade
Trade Stop Stock Indic	Text	Indicate if the trade is a Stop Stock

Trade Correlation Indic	Text	Trade Correlation Indications
Trade Sequence Number	Number	Unique sequence number for the trade

Table 2.1: Daily Trades File Table Schema

Field Name	Data Type	Description
Time	Date	Time of trade (HHMMSSXXX)
Exchange	Text	The Exchange where trade took place
Symbol	Text	Stock symbol
Bid Price	Number	Price of bid
Bid Size	Number	Bid size in units of trade
Ask Price	Number	Price of ask
Ask Size	Number	Ask size in units of trade
Quote Condition	Text	Condition of Quote issued
Bid Exchange	Text	The Exchange associated with the Bid
Ask Exchange	Text	The Exchange associated with the Ask
Sequence Number	Number	Message sequence number
Best Bid Exchange	Text	The Exchange associated with the Best Bid
Best Bid Price	Number	Best Bid price
Best Bid Size	Number	Best Bid size in units of trade
Best Offer Exchange	Text	The Exchange associated with the Best Offer
Best Offer Price	Number	Best Offer price
Best Offer Size	Number	Best Offer size in units of trade

Table 2.2: Daily NBBO File Table Schema

2.2. Data Processing

The majority of the project is to process and convert the raw database into manageable tables with basic statistics using Microsoft SQL Server for further analysis. Our initial approach to data processing was to calculate potentially analytical statistics of each stock in five-minute intervals for a selection of stocks traded both on IEX and other big Exchanges. The statistics chosen for measuring market micro characteristics for each stock are those may potentially change due to the existence of IEX. The list of statistics calculated in five-minute intervals is shown in Table 2.4.

Statistics List (5-min)
Average Price
High Price
Low Price
Price Range
Price Movement
Trade Volume
Dollar Volume
Trade Count
Volume Weighted Price
Return
Time Weighted Spread
Volume Weighted Spread
Effective Spread

Table 2.4: List of Statistics

2.3. Computational Improvement

From a programming point of view, our initial algorithm for data processing involved using cursors and loops to calculate five-minute-interval statistics for 50 stocks of the highest market capitalization. And we grouped the calculations by stocks and exchanges by looping through a Stock ID List and an Exchange List. This naïve program took about five hours to process Trade data of a single day, and more than two days to process an NBBO data, which was too slow if we wanted to analyze on the data scope of two month before and two month after the opening of IEX in limited time.

With the help of our mentor, Paca Nguyen, we revised our code by implementing the whole process using joins between tables instead of cursors and loops, so that the program only needs to read the raw data twice to finish all the calculations. After discussions about analytical methods and regression models with Paca, we also decided to group the calculations only by stocks, but not Exchanges, and we only needed the top 10 stocks listed on IEX as well to speed up the progress. In addition, we also took out some of the unnecessary columns of the raw data before the calculation to save time for table reading.

After all these adjustments, we are able to run the program and compute all the statistics of one day within two hours. And we also set an autopilot through SSIS (SQL Server Integration Service), which would automatically import, process and export statistics table from day to day.

3. Statistical Analysis

3.1. Assumptions

To detect whether IEX's establishment has an effect on the trading costs and volatility of the market, we focused on changes of various spreads, price range, price movement, trading volume and cross-trades happening. Compared with the trading volume on main exchanges, IEX, as one of the dark pools, only occupies a small part of trades in the whole market. Consequently, we firstly made the null assumption that the launch of IEX has no influence to the mean of the daily-level statistics mentioned above.

3.2. Two Sided T-Test

With the null hypothesis that there is no difference between the means of daily-level statistics concerning market spreads and volatility before IEX's opening and those after IEX's opening, we launched the two-sided t-test.

	Mean Before	Mean After	Difference of Means	95% Interval	P-value	T-value
Time Weighted Spread	0.00204776	0.00226358	-0.00021582	-0.00046134	0.00002971	0.0838 -1.7595
Volume Weighted Spread	0.04223842	0.05037804	-0.00813962	-0.01292548	-0.00335375	0.0012 -3.4044
Effective Spread	0.03100587	0.05139003	-0.02038416	-0.05012414	0.00935582	0.1753 -1.3720
Average Price	166.71890000	196.22580000	-29.50690000	-48.92445000	-10.08930000	0.0036 -3.0525
Price Range	0.24632390	0.26751580	-0.02119190	-0.05083274	0.00844901	0.1578 -1.4311
Price	0.00144342	0.00143133	0.00001208	-0.00018626	0.00021043	0.9021 0.1240

Movement							
Trade Volume	186606.1000	175823.7000	10782.4000	-7811.9760	29376.7330	0.2505	1.1607
Dollar Volume	15374438.0000	14929631.0000	444807.0000	-1900935.0000	2790550.0000	0.7056	0.3796
Cross Trade	739.3344	678.0812	61.2532	-2.2910	124.7973	0.0586	1.9295
Count							

Table3.2: Two-Sided T-Test of daily average statistics

From Table3.2 we can find that only volume-weighted spread refused the null hypothesis that we made on a 95% confident level, with p-value being 0.0012(smaller than 0.0025). And there exists a 19.27% increase in the mean of daily volume weighted spread after the IEX launching.

However, to further explore that whether volume weighted spread really increase because of IEX's establishment, we still have to get rid of various factors from the market.

3.3. Regression Models

Since the market environment such as volatility, trading volume and seasonal patterns may also affect the volume-weighted spread, we can make the linear regression of volume-weighted spread on market factors such as VIX, SP500 level, SP500 volume, etc. In this way, we can add the dummy variable of IEX to see whether this variable is significant in the linear regression model so that the effects of market environment can be eliminated. If the IEX dummy is important, then we can say that the launch of IEX has affected the spread of the market. And we made this analysis on statistics of the single stock as well as the statistics of the integral market (average of individual stocks).

For single stock, we regress the daily volume-weighted spread on the following independent variables: today's VIX, S&P500 level, S&P500 return, S&P500 volume, stock's volatility, stock's intraday return, stock's trade volume, dummy variables for Monday, Tuesday, Wednesday, Thursday, and dummy for IEX. In total, there are twelve independent variables.

We include VIX to reflect the fluctuation of the whole market. It's calculated as the

average of today's VIX open and close. We include S&P500 level, return and volume to reflect the temperature of the market. The S&P500 level equals the average of todays' open and today's adjusted close. Today's S&P500 return equals the natural log of today's adjusted close divided by yesterday's close. The data of both VIX and S&P500 is imported from the .csv file we downloaded from Yahoo finance.

We include stock's daily volatility, return, and trade volume to reflect the market trend on the single stock. The volatility of the stock is calculated as the standard deviation of volume-weighted price of five minute's interval during trading time. The intraday return is calculated as natural log of today's close divided by today's open of the stock.

We introduced dummy variables for Monday, Tuesday, Wednesday and Thursday to reflect the pattern of the spread throughout weekdays. We expect Monday to be more significant than others, since trades on Monday reflect information of three days: Saturday, Sunday and Monday.

We introduced IEX dummy. All trading day after 2013/10/25 will be marked as 1 and days before that day will be marked as 0.

For spread on whole market, the independent factors we applied are today's VIX, S&P500 level, S&P500 return, S&P500 volume, dummy variables for Monday, Tuesday, Wednesday, Thursday, and dummy for IEX.

3.4. Presentation of Models

We regressed on data of 60 days: 31 days before, and 29 days after the launch of IEX (2013/10/23). The significance of each factor is shown below:

Significance	AAPL	XOM	GOOG	MSFT	JNJ	WFC	GE	WMT	CVX	JPM
Intercept	0.564941	0.548406	<u>0.002327</u>	0.1549	0.36497	<u>0.00909</u>	0.6977	0.94143	0.662140	0.172415
VIX	0.034335	0.860614	<u>0.000848</u>	0.6662	0.05304	0.03564	0.6288	0.43418	0.048246	0.139420

SP500 Level	0.626427	0.311193	<u>0.002263</u>	0.5757	0.11191	0.54302	0.0434	0.24623	0.961193	0.457570
SP500 Return	0.931464	0.992465	0.096111	0.9026	0.25853	0.80256	0.8356	0.55444	0.154447	0.704365
SP500Volume	0.851861	<u>0.000355</u>	<u>0.001060</u>	<u>0.0195</u>	<u>0.00483</u>	<u>0.01318</u>	0.0548	0.19347	0.067915	<u>0.000643</u>
Stock Volatility	0.751825	0.108025	0.114261	0.0918	0.21535	0.08319	0.4949	0.76986	0.237934	0.068116
Stock Return	<u>8.32e-05</u>	0.103487	<u>0.002568</u>	0.0766	0.23917	0.08330	0.4364	0.70731	0.212596	0.702652
Stock Volume	0.337152	0.474680	<u>0.011457</u>	0.6173	0.29724	0.12772	0.2611	0.25520	<u>0.004696</u>	0.025130
Monday dummy	<u>0.016487</u>	<u>0.006242</u>	0.037790	0.0864	<u>0.01557</u>	0.31940	0.6908	<u>0.01712</u>	0.763465	<u>0.003143</u>
Tuesday dummy	<u>0.000372</u>	0.417814	0.401708	0.8074	0.23919	0.33923	0.9138	<u>0.00749</u>	0.721767	0.191780
Wednesday dummy	<u>0.001246</u>	0.389204	<u>0.001493</u>	0.1811	0.35179	0.76982	0.8018	0.05989	0.873083	0.035916
Thursday dummy	0.032552	0.417752	0.026415	0.1778	0.10886	0.86822	0.8794	0.07922	0.147599	0.443402
IEX dummy	0.823013	0.996211	0.913933	0.7415	0.75858	0.90143	0.1420	0.64427	<u>0.000787</u>	0.463253
	AAPL	XOM	GOOG	MSFT	JNJ	WFC	GE	WMT	CVX	JPM
R-squared	0.6056	0.4327	0.6726	0.3241	0.3836	0.2942	0.2274	0.3861	0.5611	0.5038

Table 3.4: p-value of coefficients for single stock's linear regression
Note: The cells with significance level 0.05 are underlined.

Then we do linear regression on volume-weighted spread of the whole market. The dependent variable, the volume-weighted spread, is the arithmetic average of the volume-weighted spread of the ten stocks.

$$\begin{aligned}
VSP &= 0.008735 - 0.000117382 VIX - 0.000008 SP_Return - 0.000275944 SP_Volume \\
&\quad (65.53) \quad (-1.69) \quad (-0.13) \quad (-4.58) \\
&- 0.00051 Mon - 0.00021 Tue - 0.0003 Wed - 0.00013 Thur + 0.000116 IEX \\
&\quad (-2.86) \quad (-1.13) \quad (-1.78) \quad (-0.70) \quad (0.83) \\
Df &= 51, R^2 = 0.4274
\end{aligned}$$

VSP: volume-weighted spread

VIX: centralized VIX index

SP_Return: daily return of S&P 500 index

SP_Volume: volume of the S&P 500 index

Mon: dummy variable indicating whether it is Monday

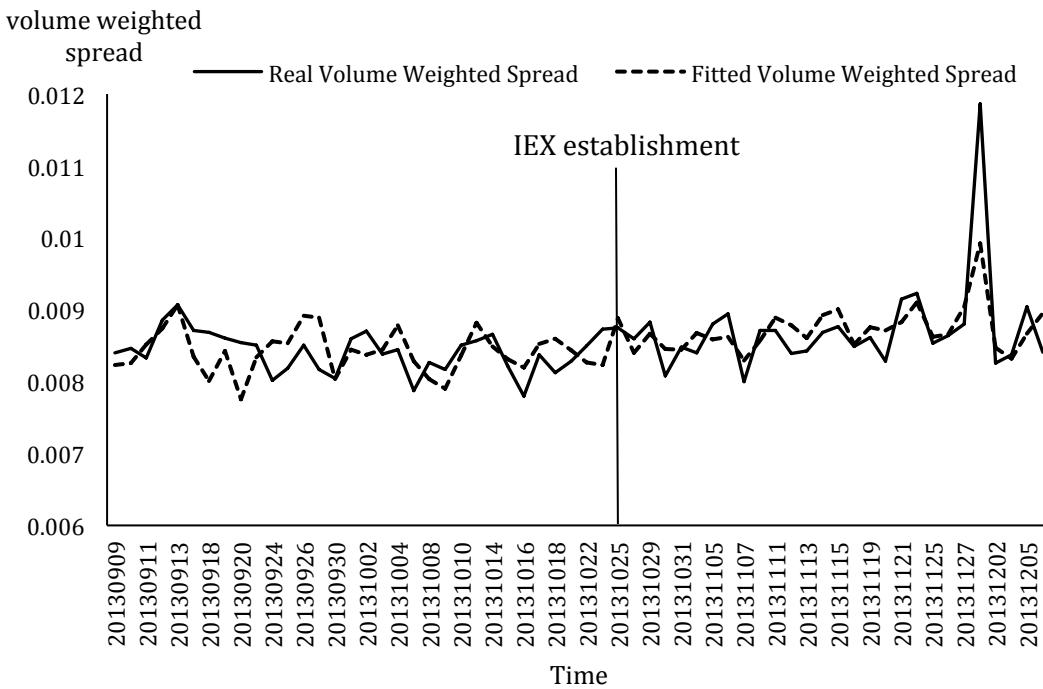
Tue: dummy variable indicating whether it is Tuesday

Wen: dummy variable indicating whether it is Wednesday

Thurs: dummy variable indicating whether it is Thursday

IEX: dummy variable indicating whether it is before or after IEX establishment

(0-before, 1-after)



From the regression above Monday effect and market volume has a great influence on the volume weighted spread, while IEX's effect is not significant after taking the market environment into consideration. The t-value is 0.83, on 95% confident level; we can say that there is no significant difference of volume-weighted spread before and after IEX establishment after eliminating the trend of the whole market.

4. Summary

Based on the statistical analysis, we couldn't find any significance on transaction cost (which is denoted as "Spread") and volatility (denoted as "Price Range"), etc.

This might because that the market share of IEX is too small or because the data range

that we used is too short for those effects to be significant. There do exist possibility that if we make the data range longer, we might find out some significances. Unfortunately, we have time limitation to process more data.

In spite of these findings, we still believe that IEX is a great revolution in the market. We look forward to seeing more benefits that IEX will bring to the market.

6. Reference

- Hope, Bradley. “IEX Pricing Aims to Drain ‘Dark Pools’”. The Wall Street Journal. 6 July. 2014. Web. 10 Nov. 2014. <http://online.wsj.com/news/article_email/iex-pricing-aims-to-drain-dark-pools-1404687705-lMyQjAxMTA0MDAwMzEwNDMyWj?cb=logged0.4440955277532339>.
- Chakrabarty, Bidisha. Moulton, Pamela. Shkilko, Andriy. “Short sales, long sales, and the Lee-Ready trade classification algorithm revisited.” Journal of Financial Markets 15 (2012): 467-491. Web. 1 Sep. 2014.
- Securities and Exchanges Commission, 2007. “Economic Analysis of the Short Sale Price Restrictions Under the Regulation SHO Pilot.” Retrieved on 1 Sep. 2014.
- NYSE Euronext, 2014. “Daily TAQ Client Specification.” Retrieved on 10 Sep. 2014.