SEC Rulings on Dark Pool Transparency: A Case Study to Understand Its Effects on Price Discovery

Srivats Narasimhan and Eric Zhang
Research Mentor: Mark Flood

Srivats Narasimhan is a SURE Fellow in the Department of Computer Science and Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: snarasim@terpmail.umd.edu.

Eric Zhang is a SURE Fellow in the Department of Computer Science and Robert H. Smith School of Business at the University of Maryland, College Park, MD. Their email address is: ericzhang98@gmail.com.

Abstract

This paper examines recent SEC rulings that demand increased transparency in Dark Pools and their effects on price discovery, using Cisco as a case study. A two-week and a four-week sample of two different liquidity measurements, Amihud's measurement and Turnover ratio, were taken a month before the SEC rulings and a month after the implementation of the SEC rulings. These samples were converted to time-series, and a Granger Causality test was then performed on each separate group of time-series. Four of the five Causality tests yield the same result of not causal. The fifth result yielded a causal relationship; however, this anomaly could possibly be attributed to exogenous variables in the market. Because there is no causal relationship present on a general level, based on the results of this study, it can be theorized that the SEC rulings did not affect price discovery. Yet, due to limitations of the case study itself, its results may not entirely indicate the implications of the SEC rulings on price discovery at a larger, more generalized level.

Keywords: Dark Pools, Granger Causality, Time series, Amihud's Measure, Turnover Ratio, Down sampling, Java, R

Introduction

Dark pools have played a pivotal role in the financial markets in recent years. In fact, Rosenblatt Securities estimates that dark pools execute 14% of the U.S. equity volume (*Bain*). According to the CFA, Dark pools have surged in usage, accounting for an estimated 40% of all US stock trades in Spring 2017, compared to an estimated 16% in 2010 (*Dark Pools)*.

The origin of Dark Pools can be traced back to the 1980's. However, the recent book *Flash Boys,* written by Michael Lewis, has given dark pools public attention. The book described how many proprietary trading firms used tactics, such as 'pinging' dark pools, to unearth large orders that were hidden, then engage in front-running and latency arbitrage to generate Alpha. This poses the following questions: What exactly are dark pools, and how do they impact financial markets? Although the term may colloquially sound rather ominous, it refers to a standard market practice. Dark Pools, or Alternative Trading Systems (ATS), are private exchanges intended for the trading of securities and are usually operated by larger financial institutions. It is important to note that Dark Pools are not readily accessible to the general public, but only to institutional investors with large amounts of capital. The primary facet of Dark Pools is the lack of transparency for general traders in the LIT markets. Typically, dark pools have limited or no pre-trade transparency, anonymity, and mid-quote pricing (*Buti, Sabrina and Rindi, Barbara*). The premise of Dark Pools is to prevent large amounts of orders from affecting the price discovery of certain assets in LIT markets.

The debate on the morality and legality of Dark Pools has sparked a recent general interest in whether they actually affect equity price discovery in the LIT markets. The theoretical premise of dark pools is to prevent the manipulation and saturation of LIT markets from affecting returns for institutional investors that typically make significant trades. However, as of July 18th, 2018 (*Bain*), the SEC ruled to increase transparency of dark pools. The ruling came into effect on October 9th, 2018.

The purpose of this paper is to investigate the SEC rulings to increase the transparency of Dark Pools and their potential impact on the price discovery and efficiency of common equities. The SEC policy has tremendous implications on the financial industry, potentially enabling an increase in the amount of informed trades in LIT markets. The very premise of Dark Pools allowed investors to make informed trades without having to deal with the related implications that information would have on the general market. With recent SEC rulings, however, it is important to understand whether price discovery and efficiency actually increase for equities, because of newly disseminated information that previously wouldn't have reached LIT markets.

This paper is an exploratory case study of Cisco and its stock value. To explore the possible impact of the ruling, this paper used two separate liquidity measures: The Turnover Measure and Amihud's Measure. Time-series of these measures were generated, and Granger Causality tests were run to determine if there is a causal relationship between the time-series before and after the implementation of the discussed SEC rulings. The time frame of this study took place within the past year, since when the SEC rulings were implemented. This paper hopes to provide initial research into the impact of the SEC rulings on equities and to provide some insights into the debate how dark pools actually impact LIT markets.

Case Study: CISCO

This paper represents an exploratory case study, and is not to be misunderstood as an authoritative econometric analysis. As such, analyzing one stock would effectively explore the paper's intended goal, to understand how new SEC rulings on Dark Pools affect price discovery and efficiency. It would be increasingly more difficult to analyze price discovery and efficiency for more than one stock.

Certain criteria were essential in selecting a stock for this study. The stability of the firm itself is important, because it ensures there will be fewer idiosyncratic risks associated with small and high growth firms. These risks usually include high volatility in share price. The main measure of volatility in a stock is the Beta Coefficient, which represents how the share's returns relate to those of the general market. The second important criterion is the nature of the firm itself. Choosing a larger, more established conglomerate would lessen the risk associated with highly specialized, single-industry firms. The third, and final, criterion is that the stock is traded on the S&P 500, to confirm the firm's established nature and readily available data.

Based on these three criteria, the study chose to focus on Cisco (Ticker Symbol: CSCO). The Beta Coefficient of Cisco currently hovers around 1.06 according to Yahoo Finance; this implies that Cisco is a relatively stable firm whose returns tend to reflect that of the general market (i.e. it tends to move with the market). This is a relatively low Beta Coefficient, given the fact that Cisco operates within the software and networking hardware industry. However, because Cisco is a conglomerate and is diversified within the Technology Industry, it is less prone to concentration risk than its related counterparts and competitors. Finally, Cisco is traded on the S&P 500, indicating that it is established and that data is readily available for retrieval and corresponding analysis.

Methodology

The basis of the analysis relies on high-frequency data. High-frequency data elucidates many essential measures to help understand what impacts price discovery. The main function of this data collection is to construct the time-series that are to be used in the study's Granger Causality tests. The source of this data is the Wharton Research Data Services (WRDS) database; from there, access to the Trade and Quote Database (TAQ) provided relevant observations. It contains intraday transactions regarding trading and quote data for all securities exchanges, at a millisecond scale. The SEC rulings concerning increased transparency in dark pools were announced on July 18, 2018, and implemented on October 8, 2018. The studies' data points were extracted from two intervals. The first type tested intervals that roughly spanned two weeks – those that took place approximately a month before the SEC rulings were announced and one month after the SEC ruling rulings went into effect. The second type of tested intervals spanned roughly four weeks -- these took place approximately one month before the SEC rulings were announced and a month after the SEC ruling rulings went into effect. The two-week span taken before the SEC rulings occurred from June 4 to June 18, 2018. The four-week span taken before the SEC rulings occurred from May 21 to June 18, 2018. The two-week span taken after the SEC rulings occurred was

November 5 to November 19, 2018. The four-week span taken after the SEC rulings occurred was October 22 to November 19, 2018.

Down-sampling

Down-sampling is critical to the functionality and analysis of the dataset. It is typically done to conserve memory from a computing standpoint and to reduce signal processing time when data has been oversampled. The high-frequency data that was utilized in this paper comes with data-points per millisecond. This study’s most relevant statistic concerns the volume-traded-per day. Down-sampling is the action of aggregating large data-sets into smaller, more focused intervals. For this specific case study, the data was down-sampled from a high frequency (data-per-millisecond) into a lower frequency (data-per-day). Through the usage of code written in Java, the data-per-millisecond was parsed and stored into a data-structure that contained daily totals in trading volume. This was then used to compute the liquidity measures and construct the corresponding time-series.

Granger Causality

A time-series is basically a set of quantitative observations arranged in chronological order. Specifically, with time-series, time is a discrete variable, meaning that when a time-series analysis is run, only a certain time interval can be allocated for data-analysis. Usually, time-series help measure the forecast error, which is the difference between an observed value and its forecasted value. This study's time-series consist of the measurements of liquidity (derived from high-frequency data) from the weeks specified above.

The specific analysis within our time-series is Granger Causality, which is a test of two different time-series used to determine the relationship between each time-series. The premise of a Granger Causality test is to determine if one time-series is causal for another time-series; in other words, a Granger Causality test helps show if one time-series can predict another.

From here, two separate time-series, X and Y, will be examined. Let I_t represent the the specific data available at a certain time -- t. When translating this in terms of the whole time-series, X_t is, therefore, representative of the current and past data available for the time-series X at time t. The same would apply for Y, where Y_t would, therefore, be representative of the current and past values available for time-series Y at time t. When translating into a set equation,

$\underline{X_t}$ would be equal to the set of all current and past values of X, where $\underline{X_t}$= $\{X_t, X_{t-1}, X_{t-K}\}$, and $\underline{Y_t}$= Set of all current and past values of X, $\underline{Y_t}$= $\{Y_t, Y_{t-1}, Y_{t-K}\}$. The corresponding forecast error is denoted as σ^2.

There are explicit two separate measurements of Granger Causality: (1) Granger Causality and (2) Instantaneous Granger Causality. The following literature provides explanations of the two separate measurements:

(1) X is Granger Causal to Y if the following equation is satisfied:

$$\sigma^2(y_{t+1} | I_t) < \sigma^2(y_{t+1} | I_t - \underline{X_t}).$$

In this equation, $\underline{X_t}$ can be described as the lag data set, and the equation y_{t+1} is used to denote the current time-series data set. This equation suggests that the future values of the time series Y, denoted by y_{t+1}, can be better predicted if the current and past values of X are used.

(2) X is instantaneous Granger Casual to Y if the following equation is satisfied:

$$\sigma^2(y_{t+1} | \{I_t, X_{t+1}\}) < \sigma^2(y_{t+1} | I_t).$$

This equation states that the future value of Y, denoted by y_{t+1}, can be better predicted if the current and past values of X, denoted by X_t, and if the future values of the time series X, denoted by X_{t+1}, are used within the equation.

Instantaneous Granger Causality is a part of the Granger Causality theoretical framework; however, this paper will not test for Instantaneous Granger Causality and will focus on Granger Causality.

Liquidity Measures

One of the most debated topics regarding the effects of dark pools on LIT markets involves the concept of liquidity. Liquidity can be understood as the matching of the demand and supply of a given asset with the least associated transaction costs (*Gabrielsen, Marzo, and Zagaglia*). More simply put, liquidity defines the idea that there exists a buyer for a seller in the market.

It is generally understood that the impact of liquidity on an asset has great implications to its pricing. From a conceptual standpoint, if an asset is illiquid, its rate of return must increase to compensate for the increased transaction costs that resulted from the illiquidity. In other words, if an investor cannot sell an asset that easily, they have to be fairly compensated as such. The Efficient Market Hypothesis (*Farma*) also dictates that, for an efficient market to exist, the market itself has to be large, liquid, and have easily accessible information for all market participants. Prices are therefore more inaccurate in illiquid markets.

Amihud's Measure

$$ILLIQ_T^i = \frac{1}{D_T}\sum_{t=1}^{D_T}\frac{|R_{t,T}^i|}{V_{t,T}^i}$$

One helpful calculation related to illiquidity is Amihud's Measure. In this case, D_T = is the number of days for which the data is actually available; $V_{t,T}^{i}$= is the daily trading volume (the dollar amount of shares traded on that specific day); $|R_{t,T}^{i}|$= is the return on that day -- t --for the given share. For this study's scope, this liquidity measure is realistically calculable because of the availability of the related data. In other words, there exists high-frequency data that actually allows the calculation of this liquidity measure, therefore making it a candidate for the time-series that were used.

To compute Amihud's Measure, several variables were factored in. The first calculation performed was the daily return of Cisco. Let X_t denote the daily closing stock price. The daily return is computed using the equation $(\frac{X_{t+1}}{X_t})$ - 1. The next variable that was essential to the calculation is the daily dollar value. This measure is simply computed by multiplying the daily stock price by the daily volume traded. Finally, Amihud's measure is computed by taking the absolute value of the daily return and dividing it by the daily dollar value.

Turnover Measure

$$TR_t^i = \frac{Sh_t^i}{NSh_t^i}$$

The turnover measure, or ratio, is the second liquidity measure that was utilized in the case study. Essentially, because the equation entails the average amount of shares traded over the total amount of shares outstanding (available to the market), it can act as a measure of how difficult it is to sell a share in the market. As such, time-series were created based on turnover measure calculations.

In order to compute the turnover measure, only two specific variables are needed. Let $NSh\ _t^i$ = the volume traded on a specific day. Let $Sh\ _i^t$ = total amount of Cisco shares outstanding. The latter was found within Cisco's quarterly reports and was used as a constant for the calculation.

Each liquidity measure was calculated and compiled into a Bivariate time-series model.

Before the Granger Causality test was run, the null and alternate hypotheses were established based on time-series X: pPre SEC ruling, and time-series Y: post-SEC ruling.

- Null Hypothesis = X **does not** Granger Cause Y
- Alternative Hypothesis = X **does** Granger Cause Y
- If the resulting p value < 0.05, the Null Hypothesis is to be rejected. Rejecting the Null shows that X does Granger Cause Y, revealing that the values of Y can be predicted by X.
- If the resulting p value > 0.05, the Null Hypothesis is to be accepted. Accepting the Null hypothesis shows that X does not Granger Cause Y, revealing that the X cannot predict the values of Y.

Also, one lag variable was chosen when performing the Granger Causality test. One variable was ideal because of the limited sample size, where too many lag variables would have decreased the reliability of the results. If Granger Causality emerges after regulations, it says that informative trades are migrating into dark pools, and LIT markets are less informed, only learning what they learn after dark pools have completed their trades. Thus, if pre-SEC ruling liquidity measures Granger Causes post-SEC ruling liquidity measures, it can be said for certain that the SEC ruling does impact price discovery.

Limitations

There were a few key limitations to the study's research scope. The main limitation involved lack of a substantial sample size. To perform an efficient and more reliable Granger Causality test, years and years of observations -- in other words, data-points -- are typically needed. However, for this study retrieving this data was not possible. The study was able to retrieve stock-related data for only two-week and four-week intervals before and after the implementation of the SEC rulings. Regardless of the fact that a Granger Causality test may not be as effective in cases with smaller data-sets, it doesn't necessarily discount the resulting conclusions.

The team decided to approach this research project as more of a case study, mainly because there were too many potential data points if we considered more than one specific equity. Because this study was completed over the course of only one semester, the data had to be reasonably sized. Personal computers were used to conduct the data analysis, and these devices did not possess the required computing power to handle large data-sets pulled from a general index (many stocks). All of these limitations led to framing the case study to analyze just one stock. After initial background research, it was found that Cisco (CSCO) could provide a valid sample because it is a stable stock, with a Beta Coefficient close to 1; this meant that general conclusions could potentially represent price discovery on a higher level as well.

Another consideration that must not be forgotten is that investors are also humans, that a psychological factor is involved in the investment process. In a sense, transitive and anomalous behaviors might occur as a result

of the SEC rulings, which may affect the resulting data-sets that were utilized in the study.

Data Analysis

Down-Sampling Code

Performed in Java

Working with high-frequency data is difficult, due to the sheer amount of data points that must be manipulated. Even with just one share's (CISCO) high-frequency data, down-sampling had to be performed to create time-series for each individual liquidity measure that was used in our Granger Causality tests.

To actually down sample the code, Java's Excel API, Apache POI, was used to parse through the high-frequency data (which were stored in CSV files). A Linked-List data-structure was used to create individual nodes that stored trade volume on a specific day; a Linked-List was used because of its dynamic nature and ability to store data within each individual node. Because the data was pulled in millisecond intervals, it would have been ineffective and inefficient to use each trade-volume-per-millisecond metric as part of the time-series.

Data Analysis Code

Performed in R (or RStudio)

Essentially, in order to actually perform Granger Causality tests for the discussed liquidity measures, some sort of statistical package capable of performing it natively was required. Creating the actual computing backend of this statistical analysis is unnecessary for the purpose of this study. Thus, after using the down-sampled data, each individual liquidity measure was calculated in separate spreadsheets (through formulas in Excel), then imported the respective data into RStudio. There, using Granger Causality functions (found in a package downloaded for RStudio), the statistical analysis was performed.

Results

Amihud's Measure

After the data for the Amihud's Measure was computed in an Excel spreadsheet, the two-week periods before and after the SEC ruling were compiled into a Bivariate time-series model.

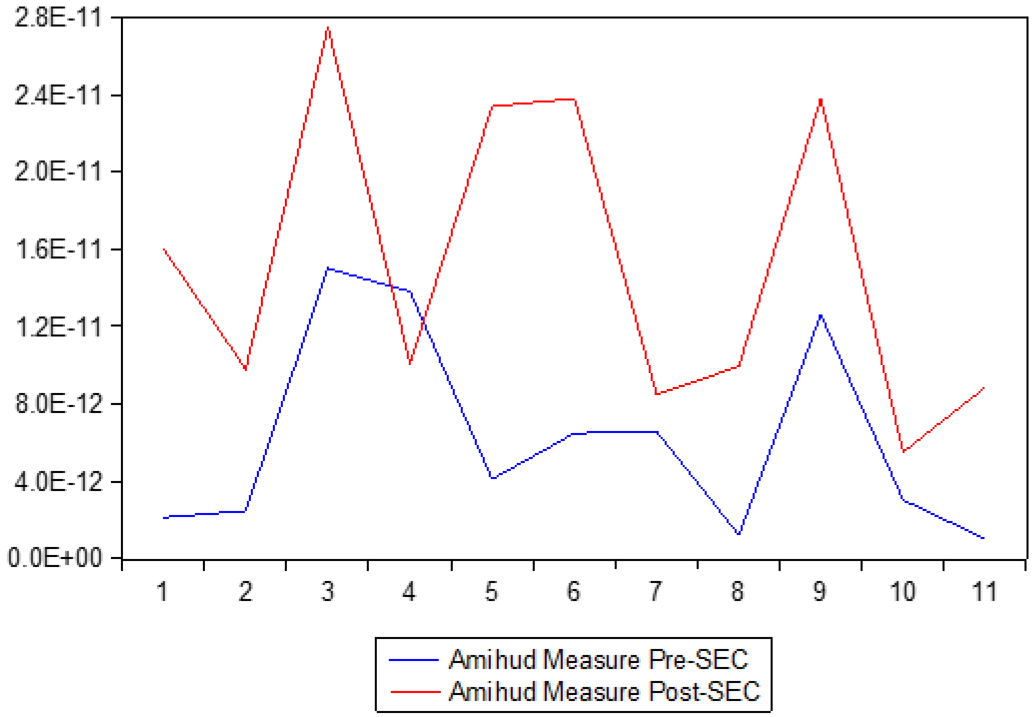

Fig 1: Time Series of Amihud's measure from the two-week interval

As seen above, the average Amihud Measure pre-SEC rulings was 0.0000000000062227, and the average Amihud Measure post-SEC rulings was 0.0000000000152036. It can be noted that within this two-week period, the post-SEC rulings' Amihud Measure nearly doubles its older counterpart. Subsequently, this can be visually noted in the time-series detailed above.

The Granger Causality test based on the two-week sample of the Amihud Measure was run and yielded the following results:

Res.	Df	Df	F-value	Pr (>F)
1	7			
2	8	-1	0.192	0.6745

Fig. 2 Granger Causality results for Amihud's Measure for the two-week interval

The following test revealed a P-Value of 0.6745. Since the P-value is greater than the 0.05 threshold already established, the Null Hypothesis could not be rejected. This reveals that within this two-week span, the SEC rulings were not Granger Causal.

The data was then expanded to include a larger sample size of 4 weeks. This increase in sample size was conducted to determine the consistency of the previous results and to reduce the effect of any exogenous variables that might affect the Amihud Measure. This data was then converted into another Bivariate time-series model.

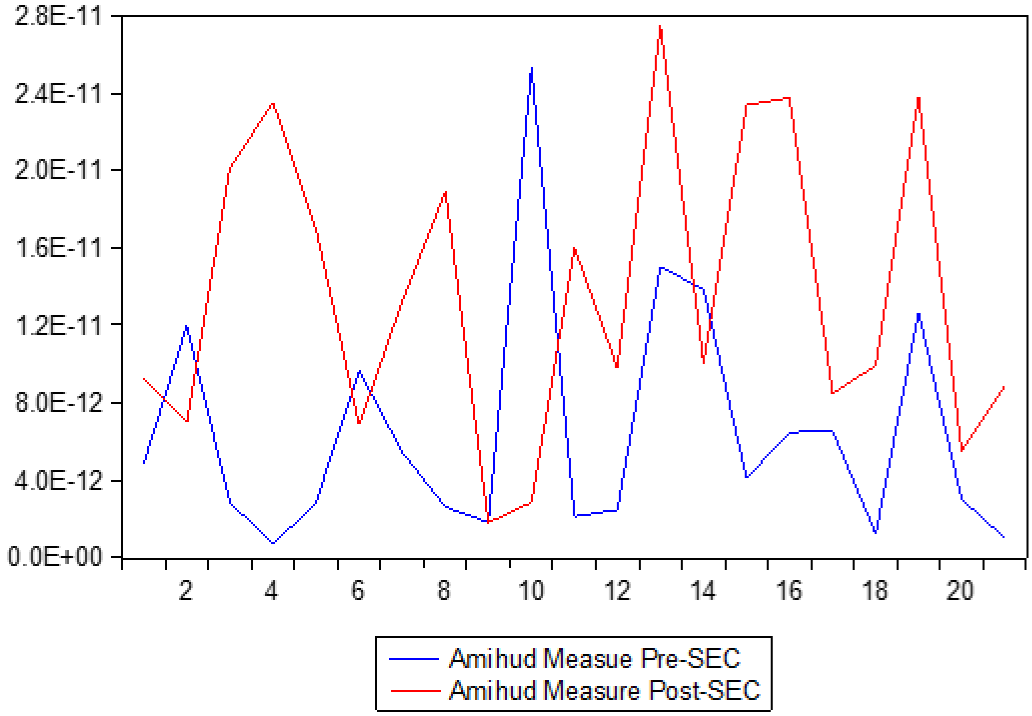

Fig 3: Time Series of Amihud's Measure from the four-week interval

Through a brief visual analysis of the previous model, there are a few important ideas to note. Both time-series appear to be relatively volatile, and, akin to the previous observation, it seems that there is an increase in the Amihud Measure after the SEC rulings. The average pre-SEC rulings value was 0.0000000000065084, and the average post-SEC rulings value was 0.0000000000136988. The same doubling effect observed in the two-week sample can also be observed in the four-week sample.

The Granger Causality was then run on the four-week sample of the Amihud Measure and yielded the following results:

Res.	Df	Df	F	Pr (>F)
1	17			
2	18	-1	0.0193	0.8912

Fig. 4 Granger Causality results for Amihud's Measure for the four-week interval

This following data resulted in a p value > 0.05, thus leading us to accept the Null Hypothesis. This increase in sample size revealed the same conclusion as the two-week sample size, that the SEC ruling was not Granger Causal. Because the sample size was effectively doubled and the same result was achieved, a level of consistency is shown.

Both the time-series of the two-week sample period and the four-week sample period failed to reject the Null Hypothesis. This allows one to conclude that, in terms of Amihud's Measure, there is no Granger Causality. The values of the pre-SEC ruling Amihud's Measures do not Granger Cause post-SEC ruling Amihud's Measures. Based solely on this, it seems that the SEC rulings did not affect price discovery and efficiency. To further justify these findings, the Turnover Measure was then used.

Turnover Measure

The Turnover Measure was the next liquidity measure to be considered in this study. Similar to Amihud's Measure, the Turnover Measure was computed for before and after the SEC rulings. This liquidity measure was compiled over two-week and four-week interval, and was constructed into a Bivariate time-series.

The first two-week time-series was constructed based on the Turnover Measure, as detailed below:

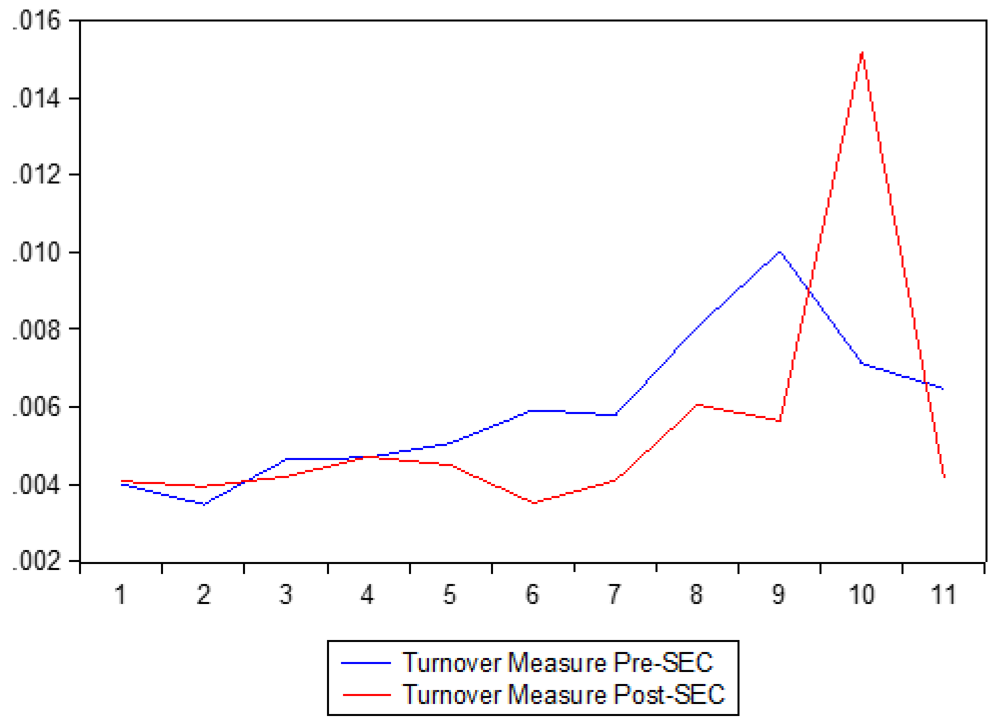

Fig 5: Time Series of Turnover Measure from the first two-week interval

The initial observations reveal an interesting trend between the two time-series. Both time-series seem to follow a similar trend, wherein they both increase significantly and decrease around the same time. The average Turnover Measure pre-SSEC rulings was 0.005935654, and the average Turnover Measure post-SEC rulings was 0.005461167. Contrary to Amihud's measure, the average turnover measure was lower post-SEC ruling. When observing the graph, the pre-SEC ruling measure outperforms the post-SEC ruling except on day 10, when there seems to be a spike in trading in the time sample post-SEC ruling. This spike p could potentially skew the average and the causality test.

A Granger Causality test was then performed on the time-series and yielded the following result:

Res.	Df	Df	F	Pr (>F)
1	7			
2	8	-1	17.252	0.004276

Fig. 6 Granger Causality results for Turnover Measure for the first two-week interval

The causality test revealed a p value of 0.004. This p value is < 0.05, meaning that the Null Hypothesis is to be rejected, and the Alternative Hypothesis would be accepted. Based on the results of the causality test for the turnover liquidity measurement over two weeks, the post-SEC rulings time-series is casual to the pre-SEC rulings, and there is Granger Causality. This result seemed strange at first, because it deviates far from Amihud's Measure.

To confirm this causal relationship, two alternative tests were run. A different two-week period, October 22 to November 4, 2018, was sampled. The turnover for this two-week period was therefore calculated, and a different time-series was generated; this is detailed below:

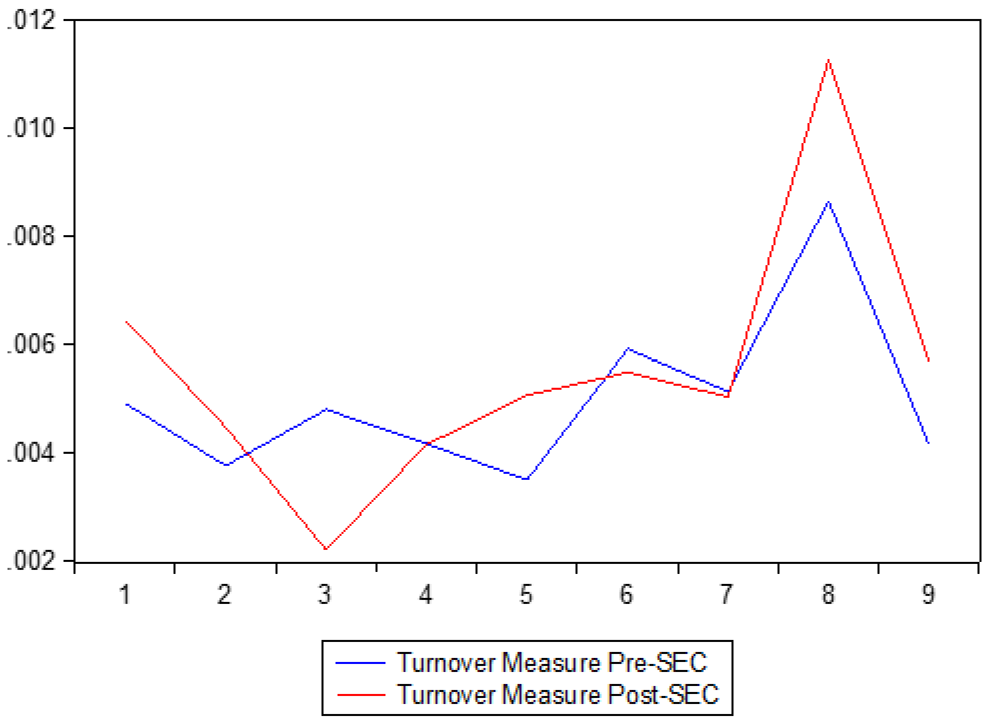

Fig 7: Time Series of Turnover Measure from the second two-week interval

With this different two-week sample size, the average Turnover Measure pre-SEC ruling is 0.005006049, and the average Turnover Measure post-SEC ruling is 0.005537827. A slightly higher average in Turnover Measure is noted in the graph as well.

After conducting another Granger Causality test, the following results were yielded:

Res.	Df	Df	F	Pr (>F)
1	5			
2	6	-1	0.1536	0.7112

Fig. 8 Granger Causality results for Turnover Measure for the second two-week interval

The following revealed a p-value that is greater than 0.05 which leads to the conclusion that the pre-SEC rulings time-series does not Granger Cause post-SEC rulings time-series. This alternative result challenges the previous two-week interval time-series, which indicated a causal relationship.

A third test involving a four-week period sample size was tested, and the following time-series was constructed:

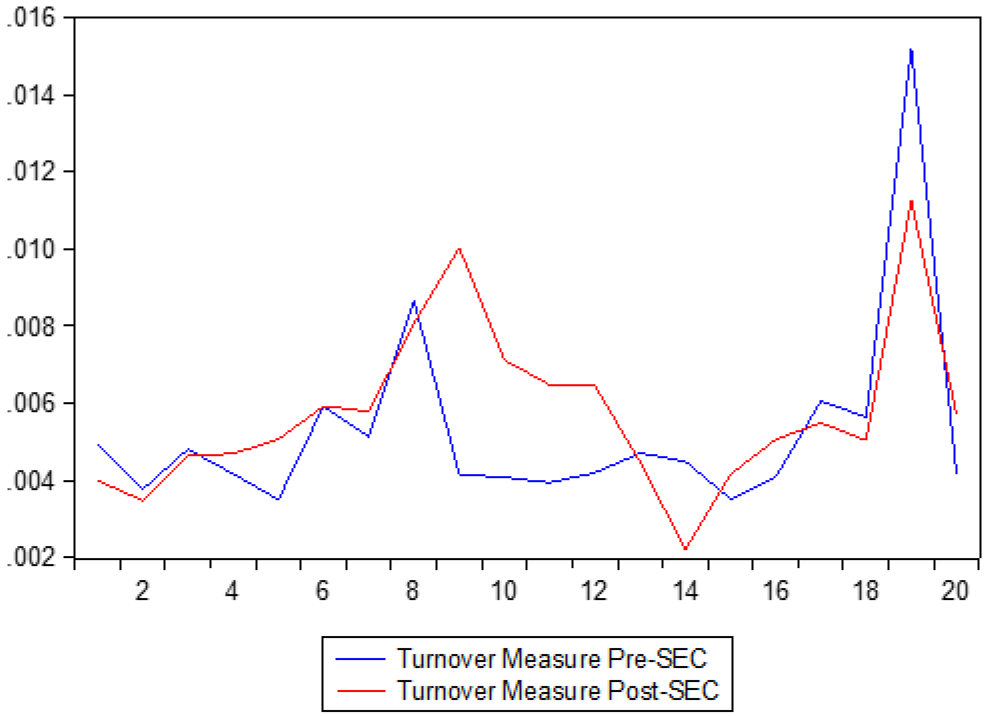

Fig 9: Time Series of Turnover Measure from the four-week interval

The Turnover measure pre-SEC ruling within this 4-week sample size was 0.005256364, and the average Turnover measure post-SEC ruling was 0.005756632. There is also an increase in the average turnover measure within this 4-week period.

The Granger Causality test yielded the following results:

Res.	Df	Df	F	Pr (>F)
1	16			

2	17	-1	0.0577	0.8132

Fig. 10 Granger Causality results for Turnover Measure for the four-week interval

The p-value for the four-week Turnover Measurement revealed a p value > 0.05, which leads to a failure to reject the Null Hypothesis. Based on the four-week data, this yet again leads us to the conclusion that pre-SEC ruling does not Granger Cause post-SEC ruling.

Although the first two-week test concluded a Granger Causal relationship, the next two-week test refuted this claim. From this we can conclude that, in terms of Turnover measurement, it is highly likely that there is no Granger Causality. The first test that yielded a Granger Causal relationship may be caused to exogenous variables in the market that could have temporarily affected the results of the Causality test.

Conclusion

After conducting Granger Causality tests for our two chosen liquidity measures, Amihud’s Measure and Turnover Measure, we are able to see that, on the whole, the time-series before are not Granger Causal to the time-series after; essentially, this portrays that, post SEC rulings, liquidity (and therefore price discovery) for CISCO stock has not changed in a significant manner. The general purpose behind the study was to understand that if liquidity measures changed noticeably after the SEC rulings (or a Granger Causality emerged after regulations) and if informative trades were in fact migrating into dark pools, making LIT markets less informed beforehand. After the rulings, thus, a Granger Causal relationship would imply that LIT markets were more informed from decreased opacity in Dark Pools, thus increasing the price efficiency of equities.

While this may indicate a lack of impact on price discovery and efficiency, it is not necessarily a totally conclusive result. As indicated previously in the study, there are noticeable limitations in the methodology chosen for the study, as well as in the very assumption that CISCO represents LIT markets in general. Many exogenous variables may have affected the outcome of the study, including investor decision-making (a psychological factor), algorithmic trading decisions, and a general instability in the markets.

To perhaps extend this study, there are certainly a few steps that can potentially be taken. Having a greater amount of observations is greatly beneficial to the outcome of a Granger Causality test, as the premise of the test assumes that each time-series has an extensive amount of data behind it, thus making them more representative of general trends and movements. Furthermore, testing time-series for multiple equities would create a wider representation of LIT markets, thus increasing the validity of the results. Finally, studying a greater number of liquidity measures for various equities, both before and after the SEC rulings, would further support a general conclusion; with more liquidity measures representing a general trend, a solidified conclusion is far easier to come by.

Acknowledgments

We would like to thank Dr. Mark Flood, Dr. Christina Elson, the Ed Snider Center for Enterprise and Markets, and the Robert H. Smith School of Business for supporting this research project.

Appendices

Appendix A: Java Source Code

Read Data Class

```
package Research_Project;

import java.util.*;

import java.io.File;
import java.io.FileInputStream;
import java.io.FileNotFoundException;
import java.io.FileOutputStream;
import java.io.IOException;
import java.sql.Time;
import java.util.Date;

import org.apache.poi.sl.usermodel.Sheet;
import org.apache.poi.ss.usermodel.Cell;
import org.apache.poi.ss.usermodel.Row;
import org.apache.poi.ss.usermodel.Workbook;
import org.apache.poi.ss.usermodel.WorkbookFactory;
import org.apache.poi.xssf.usermodel.XSSFSheet;
import org.apache.poi.xssf.usermodel.XSSFWorkbook;

public class Read_Data {
	private static int inputSheetNum = 0;
	private static int outputSheetNum = 0;

	public static void main(String[] args) {
		int dataIndex = 0;
		int numOfNodes = 0;
```

```
// Opens the excel sheet

try {

        // Opens the input stream for reading

        FileInputStream fopen = new
FileInputStream("C:\\Research\\Research_Data.xlsx");

        FileInputStream fwrite = new
FileInputStream("C:\\Research\\Output_Data.xlsx");

        // Loads workbook for reading

        XSSFWorkbook inputWorkbook = new XSSFWorkbook(fopen);

        // loads workbook for writing

        XSSFWorkbook outputWorkbook = new XSSFWorkbook(fwrite);

        // Opens input sheet for reading

        XSSFSheet inputSheet = inputWorkbook.getSheetAt(inputSheetNum);

        // Opens outputSheet for writing

        XSSFSheet outputSheet = outputWorkbook.createSheet("Sheet2");

        // finds the last value of the excel sheet

        int lastValue = inputSheet.getLastRowNum();

        // initializes the LinkedList with Volume_Data

        LinkedList<Volume_Data> filtered_Data = new LinkedList<>();

        // for loop to read into the data and write as well

        for (int i = 0; i < lastValue; i++) {

                // casts the date into an int and subtracts the 2018 year
measurement

                int date = (int)
inputSheet.getRow(i).getCell(0).getNumericCellValue() - 20180000;

                // finds the seconds of the data and rounds the value

                // =HOUR(A1) + MINUTE(A1)/60, converts the time to a
decimal in Excel

                // locates the trade volume executed at the specific millisecond
interval

                double volume =
inputSheet.getRow(i).getCell(1).getNumericCellValue();

                // locates the trade volume executed at the specific millisecond
interval

                // Creates a class which includes all the specific trade data

                Volume_Data trade = new Volume_Data(date, volume);
```

```
// adds if the linkedlist is empty
if (filtered_Data.size() == 0) {
	filtered_Data.add(trade);
} else {
	// temporary variable for the current trade
	Volume_Data temp =
filtered_Data.get(numOfNodes);
	// Checks if the dates's of the trade matches
	if (temp.getDate() == trade.getDate()) {
		// retrieves the current total
		double newTotal = temp.getVolume();
		// adds the new trade volume
		newTotal += trade.getVolume();
		// updates the trade volume

	filtered_Data.get(numOfNodes).setVolume(newTotal);
		// if the second of the trade is not the same, a
new linkedlist node is added
		// if the date's don't match, a new linkedlist
node is added
	} else {
		filtered_Data.add(trade);
		numOfNodes++;
	}
}
}
numOfNodes = 0;
for (Volume_Data dataPoint : filtered_Data) {
	// Creates the row for writing
	Row rows = outputSheet.createRow(numOfNodes);
	// Creates the cell for the date
	Cell dateCell = rows.createCell(0);
	// Writes the date cell value
	dateCell.setCellValue(dataPoint.getDate());
```

```
						// Creates the cell for wring the volume
						Cell volumeCell = rows.createCell(1);
						// Writes the volume cell value
						volumeCell.setCellValue(dataPoint.getVolume());
						numOfNodes++;
				}
				// /24 -> Format Cells -> General -> h:mm converts from decimal to
time
				FileOutputStream write = new
FileOutputStream("C:\\Research\\Output_Data.xlsx");
				outputWorkbook.write(write);
				inputWorkbook.close();
				outputWorkbook.close();
				System.out.println(numOfNodes);
				System.out.println("Data Done");
			} catch (FileNotFoundException e) {
				e.printStackTrace();
			} catch (IOException e) {
				e.printStackTrace();
			}
	}


}
```

Volume_Data Class

```
package Research_Project;
import java.sql.Time;
import java.util.*;

public class Volume_Data {
	private int date;
	private double volume;
	public Volume_Data(int date,  double volume) {
```

```
        this.date = date;
        this.volume = volume;
    }
    public int getDate() {
        return date;
    }
    public void setDate(int date) {
        this.date = date;
    }
    public double getVolume() {
        return volume;
    }
    public void setVolume(double volume) {
        this.volume = volume;
    }
}
```

Appendix B: RStudio Source Code

```
installed.packages("lmtest")


library(lmtest)


ah2w<-
read.csv("c:/Research/Turnover_Causality/Amihud_Causality_2Weeks.csv",
header = TRUE, sep=",")


attach(ah2w)


Amihud_Causality_2Weeks.csv<-read.csv(file.choose(), header=TRUE)
grangertest(AH_Post~AH_Pre, order = 1, data = Amihud_Causality_2Weeks.csv)


library(lmtest)


ah4w<-
read.csv("c:/Research/Turnover_Causality/Amihud_Causality_4Weeks.csv",
header = TRUE, sep=",")


attach(ah4w)


Amihud_Causality_4Weeks.csv<-read.csv(file.choose(), header=TRUE)
grangertest(AH_POST~AH_PRE, order = 1, data =
Amihud_Causality_4Weeks.csv)
installed.packages("lmtest")



library(lmtest)
```

```
to2w<-
read.csv("c:/Research/Turnover_Causality/Turnover_Causality_2Weeks.csv",
header = TRUE, sep=",")

attach(to2w)

Turnover_Causality_2Weeks.csv<-read.csv(file.choose(), header=TRUE)

grangertest(TO_Post~TO_Pre, order = 1, data =
Turnover_Causality_2Weeks.csv)

installed.packages("lmtest")

library(lmtest)

to4w<-
read.csv("c:/Research/Turnover_Causality/Turnover_Causality_4Weeks.csv",
header = TRUE, sep=",")

attach(to4w)

Turnover_Causality_4Weeks.csv<-read.csv(file.choose(), header=TRUE)

grangertest(TO_POST~TO_PRE, order = 1, data =
Turnover_Causality_4Weeks.csv)
```

References

A. Gabrielsen, M. Marzo and P. Zagaglia

Measuring market liquidity: An introductory survey

https://arxiv.org/pdf/1112.6169.pdf

Apache POI Tutorial. (2019). www.tutorialspoint.com. Retrieved 5 May 2019, from https://www.tutorialspoint.com/apache_poi/

Bain, Benjamin-*Wall Street Dark Pools to Come Out of Shadows Thanks to SEC. (2018). Bloomberg.com. Retrieved 18 July 2018, from* https://www.bloomberg.com/news/articles/2018-07-18/wall-street-dark-pools-set-to-come-out-of-shadows-thanks-to-sec

Buti, Sabrina and Rindi, Barbara and Werner, Ingrid M., Diving Into Dark Pools (November 17, 2011). Charles A. Dice Center Working Paper No. 2010-10; Fisher College of Business Working Paper No. 2010-03-010. Available at SSRN: https://ssrn.com/abstract=1630499 or http://dx.doi.org/10.2139/ssrn.1630499

Dark Pools. (2019). CFA Institute. Retrieved 13 April 2019, from https://www.cfainstitute.org/en/advocacy/issues/dark-pools

Fama, Eugene F. "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance*, vol. 25, no. 2, 1970, pp. 383–417. *JSTOR*, www.jstor.org/stable/2325486.

Farley, Ryan and Kelley, Eric K. and Puckett, Andy, Dark Trading Volume and Market Quality: A Natural Experiment (March 16, 2018). 13th Annual Mid-Atlantic Research Conference in Finance (MARC) Paper. Available at SSRN: https://ssrn.com/abstract=3088715

Gallen, U., & Berlin, F. *(2019). Introduction to Modern Time Series Analysis | SpringerLink.* Link.springer.com. Retrieved 9 March 2019, from granger.test function | R Documentation. (2019). Rdocumentation.org. Retrieved 5 May 2019, from https://www.rdocumentation.org/packages/MSBVAR/versions/0.9-2/topics/granger.test

Lewis, M. (2014). *Flash Boys: A Wall Street Revolt*. New York, NY : W.W. Norton & Company

Sciencedirect.com. (2019). Granger Causality - an overview | ScienceDirect Topics. [online] Available at: https://www.sciencedirect.com/topics/neuroscience/granger-causality [Accessed 5 May 2019].

What Are Dark Pools? - FXCM UK. (2019). FXCM UK. Retrieved 13 April 2019, from https://www.fxcm.com/uk/insights/what-are-dark-pools/