# Market Microstructure
## Module 2, 2018

## Prof. Anna A. Obizhaeva
### New Economic School
aobizhaeva@nes.ru

## Course information

**Course Website: my.nes.ru**

**Professor's Office Hours: Fri, 11:00-12:00 in room 2.10 or by appointment**

**Class Time: Tuesday, 15:30-18:45**

**Room Number: 1.14**

**TAs:**

## Course description

Like option pricing and fixed income, market microstructure has rapidly moved from the research domain of finance professors into the real world, where competition among exchanges, development of trading algorithms, and design of robust market systems all require combining the theory of market microstructure with an understanding of institutional detail of how financial markets work in practice. Liquidity, transaction costs, trading strategies, algorithmic trading, high-frequency trading, crashes, market fragmentation, circuit breakers, market design are topics of great interest to finance professors, market participants, policy makers, and sometimes even to the general public.

In the coming decade, market microstructure has potential to become one of the fastest growing fields of financial economics. All ingredients are indeed in place: Numerous topical questions still have no answers, while theorists are well equipped with game theoretic modeling tools for developing theories to answer those questions and empiricists have access to a vast amount of financial data to test those ideas.

The course will focus on U.S. markets and Russian markets. The general principles are applicable to all markets. By taking this course a student will have an opportunity to learn (1) how trading takes place in financial markets; (2) how economic theories relate to this trading; (3) how legal, regulatory, and ethical considerations shape the trading process; (4) and how data can be examined to quantify answers to questions about transaction-by-transaction trading.

This course uses a mixture of lectures and class discussion to examine various aspects of market microstructure. The course involves a mixture of finance and economics. The course will touch on numerous public policy issues, many of which are politically controversial and most of which are likely to continue to be important for years to come.  A goal of the course is to provide students with a framework within which these public policy issues can be addresses in a useful manner.

Since "microstructure data," the transaction-by-transaction records of trades as they occur, provide valuable opportunities for learning and research, the course will contain several hands-on exercises using microstructure data. Students are encouraged to use the Python software package for data analysis. The homework assignments are designed to mimic the kinds of assignments an entry-level professional employee might be asked to perform in a research oriented professional setting related to trading securities.

## Course requirements, grading, and attendance policies

The course grade will be based on the following: participation in class, regular homework assignments, in-class tests, and a final exam. Homework will be assigned every week.

*Class participation:* Class attendance and participation are required. Before each class session, students are encouraged to read the assigned material.

*In-class Tests:* At the beginning of each lecture (except the very first one), there will be short 10-minutes tests with one question about material of the previous lecture. A goal of these in-class tests is to help students not to fall behind on the material and to keep track of attendance.

*Homework assignments:* Almost each week students have to read required material and submit a short write-up with their answers on the questions *with stars*. There are five mandatory write-ups. To get extra credit students may also answer questions with no stars. Answers should be succinct and self-contained. Long write-ups are not necessarily good write-ups. Good business writing makes points in a manner that respects the reader's time. Typically, the answers to questions will not be heavily mathematical, but mathematical intuition will often be required to address some of the questions. Write-ups should be uploaded onto mynes before the <u>beginning of class, including the very first class</u>. Students should be prepared to discuss and defend the ideas in their write-ups in class. For some questions, there is no "right" or "wrong" answer, in the sense that finance and economics professors themselves are likely to disagree about the answers to the questions. Students may talk to other students about the assigned questions, but each student should prepare a write-up individually, without looking at the write-ups of other students. Each student should list on the write-up the names of the other students with whom the student discussed the write-up.

*Projects:* There will be three empirical projects. They are presented at the end of this syllabus. The purpose of the empirical analysis assignments is to enable students to learn to use software to answer empirical questions like they arise in a professional business or research environment. The write-ups should contain results and also a code.

*Final Exam:* A final exam is in-class and closed-book. One double-sided cheat-sheet is allowed.

*Grading:* Grading will be based on questions write-ups (5x3%), in-class tests (5x3%), projects (3x5%), and a final exam (50%). Missing class is strongly discouraged. Active class participation will be reflected in the final grade as well (about 5%). Case write-ups are graded on a scale with the following interpretation: 10 = *A + +*, 9 = *A+*, 8 = *A*, 7 = *A-*, 6 = *B+*, 5 = *B*, 4 = *B-*, 3 = *C*, 2 = *D*, 1 = *F*, 0 = *missing*. Most grades are in the range 5*; 6; 7;* 8. Extra credit is added with extra points. For example, a grade of 6.0 + 0.5 = 6.5 indicates a score of 6 for assigned questions plus 0.5 points for extra credit questions.

## Academic integrity policy

Cheating, plagiarism, and any other violations of academic ethics at NES are not tolerated.

## Course materials

The schedule of classes follows. "Required readings" should be completed before class. Students should come to class prepared to discuss required readings. "Optional readings" are materials students may want to read before class or after class, but they are not required. They may be the subject of class discussion, but the discussion should presume that not all students have read them. The "bedtime readings" are completely optional and include books students may want to read in the future, after the course is finished; they are included for general interest.

The **main reference book** for this course:  Hasbrouck, Joel. "Securities Trading: Principles and Procedures." (2016). This 201-page PDF file is like a textbook for this course: http://pages.stern.nyu.edu/~jhasbrou/TeachingMaterials/STPPms12a.pdf

These three books are good references for those interested in market microstructure:

Larry Harris. *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press, USA, 2002. This book contains a wealth of institutional details, descriptions of issues, and discussion of microstructure at the level of an advanced layman.

Thierry Foucault, Marco Pagano, and Ailsa Roell. *Market Liquidity: Theory, Evidence, and Policy*. Oxford University Press, 2013. This book is a good graduate level textbook. It is not necessary for the course to have it, but students can find discussions of most topics in this course.

Joel Hasbrouck. *Empirical Market Microstructure: The Institutions Economics, and Econometrics of Securities Trading*. Oxford University Press, 2007. This text is good graduate level book on empirical market microstructure. Joel Hasbrouck has an earlier, more extensive version of this book on his webpage.

Other textbooks:

• De Jong, F. and Rindi, B., 2009. *The Microstructure of Financial Markets*. Cambridge University Press.

• Pedersen, L.H., 2015. *Efficiently Inefficient: How Smart Money Invests and Market Prices are Determined*. Princeton University Press.

**Bedtime Readings:** Microstructure is fun. If you would like to learn more about finance and market microstructure, here are some good books. None of these are required readings for this course. While all of the required readings for this course are free, you may have to pay for these books.  These are light bedtime readings about markets, including ethical issues:

• Lewis, Michael, 2015. *Flash Boys: Cracking the Money Code*. Penguin Books.

• Lewis, Michael. *Liar's poker*. WW Norton & Company, 2010.

• Schwed, Fred, 1995. *Where are the Customers' Yachts? or A Good Hard Look at Wall Street*. John Wiley and Sons.

Course contents

**I.    Trading as a Game:**
- The history and logic of trading, market making
- Trading, clearing, settlement
- Exchanges as collusive cartels with fixed commission.
- Dealers markets
- The history of trading in Russia
- Treynor's model of bid-ask spread

**II.    High-Frequency Trading:**
- HFTs' strategies
- Flash Crash of May 6, 2010
- CFTC scandal and VPIN dispute
- Implementation shortfall
- Sunshine trading versus front-running, VWAP trading
- Big data in finance

**III.    Market Microstructure Theories:**
- Perfect competition: inventory models and rational expectation models
- Imperfect competition: Kyle (1985) model of price impact
- Smooth trading model of Kyle, Obizhaeva, Wang (2017)

**IV.    Market Microstructure Invariance (Kyle-Obizhaeva):**
- Invariance as implication of dimensional analysis and leverage neutrality
- Invariance as implication of equilibrium model
- Liquidity measures and pricing accuracy, tick size, fragmentation
- Open research questions

**V.    Market Crashes:**
- Past and future market crashes
- Ruble crisis of December 2014
- Financial stability, circuit breakers, price speed bumps

**VI.    FICC Markets:**
- Market microstructure of fixed income markets, currency, commodities
- Trading liquidity and funding liquidity
- Flash Rally of Oct 15, 2014
- Economics behind fixing, LIBOR and WMR manipulations

**VII.    Other Public Policy Issues and Review:**
- Price manipulation, spoofing, layering, Sarao's case
- Review

# WEEK 1 (TUESDAY, OCTOBER 30): TRADING AS A GAME

TOPICS: Brokers versus dealers; principal–agent problem; market makers; bid and ask prices; limit orders and market orders; dealer markets and organized exchanges; network externalities; limit orders and market orders; tick size; Nasdaq versus NYSE; specialist system; Buttonwood tree; fixed commissions and commission deregulation; tick size; dealer games; Christie–Schultz papers; odd-eighths controversy; Treynor model of bid-ask spread.

**REQUIRED READING:**

**\*Hasbrouck, Joel. "Securities Trading: Principles and Procedures." (2016). Part I, Chapters 1–5, pp. 1–39; Part II, Chapters 7-8, pp. 52-66; Part III, Chapters 9 and 11, pp. 68-78, 90-101.**

**\* Jack Treynor. The only game in town.** *Financial Analysts Journal*, **51(1):81–83, 1995. Reprint of W. Bagehot. The only game in town.** *Financial Analysts Journal*, **27(2):12–22, 1971. Treynor used pseudonym Walter Bagehot in original version. Walter Bagehot is the author of Walter Bagehot.** *Lombard Street: a description of the money market*. **Project Gutenberg Etext,1878. http://www.jstor.org/stable/4470784**

OPTIONAL READING:

Christie, W. G. and Schultz, P. H., Why do NASDAQ Market Makers Avoid Odd-Eighth Quotes? The Journal of Finance, 49: 1813–1840, 1994. Available at
http://www.acsu.buffalo.edu/~keechung/TEM/Journal%20Articles/Collusion%20A%201994.pdf

Christie, William G., Jeffrey H. Harris, and Paul H. Schultz. "Why Did NASDAQ Market Makers Stop Avoiding Odd-Eighth Quotes?." *Journal of Finance 49*: 1841-1860, 1994. Available at
http://www.acsu.buffalo.edu/~keechung/TEM/Journal%20Articles/Collusion%20B%201994.pdf

Bernanke, Ben, Clearing and Settlement during the Crash, 1990, The Review of Financial Studies 3 (1): 133-151, 1990.

OPTIONAL VIDEO: AFA's interview of Jack Treynor
www.youtube.com/watch?list=PL42MdOODnBSv2CQ0T4KJXBIRdZiWPIUYK&v=TdMf9nkKWhY

Movies on the history of trading:
www.youtube.com/watch?v=tEbZswUEnsU
www.youtube.com/watch?v=--H8SY334Zw

## Questions for discussion:

1. What is the difference between a "limit order," a "market order," and a "marketable order"?

2. What is the difference between "time priority" and "price priority"?

3. What is the difference between a "broker" and a "dealer"?

4. What is the difference between a "principal" and an "agent"? What is a "principal–agent" relationship? In securities trading, who is the principal and who is the agent?

5. Why do exchanges with "floor markets" require all trading to be on the trading floor, conducted by "open outcry"? Does the open outcry system provide time and price priority to orders?

6. What does it mean to "walk the book"?

7. What does it mean take and make the market ?

8. In the days of floor trading on monopolistic exchanges with fixed commissions, why did exchange officials pay great attention to auditing brokerage firm error accounts?

9. In the days of floor markets, why were entry-level employees who carried filled and unfilled orders around the trading floor called "runners," even though the rules of the exchanged required them to walk and not run?

10. In fixing the many unintentional errors which occur with human trading, which errors are most important to fix first: (1) price errors, (2) quantity errors, (3) trade direction (buy or sell) errors?

11. Before the invention of electricity, why would exchanges light a torch or candle, then officially end trading at the moment the torch or candle went out after burning its last remaining fuel?

12. What does pre-trade and post-trade transparency ?

13. What is the difference between being "picked off" and "run over"?

14. How does a trader tell the difference between a dealer's "indicative" quoted bid-ask spread and "actionable" quoted bid-ask spread?

15. What is the difference between a "dark pool" and a "lit market"? What is the economic and regulatory logic for the existence of dark pools?

16. How does Jack Treynor explain where bid-ask spread comes from?

# WEEK 2 (TUESDAY, NOVEMBER 6): HIGH-FREQUENCY TRADING

TOPICS: Strategies of high frequency traders; smooth trading; realized volatility; Flash crash; trading halts; price limits; big data in finance; market making.

**REQUIRED READING:**

**\*Hasbrouck, Joel. "Securities Trading: Principles and Procedures." (2016). Part IV, Chapter 13, pp. 111–116 and Part V, Chapter 19, pp. 156-168.**

**\*Lawrence Harris, What to do about high-frequency trading? Financial Analysts Journal. 2013**

**\*Report of the staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues, Findings Regarding the Markets Events of May 6, 2010, September 30, 2010**

OPTIONAL READING:

Andrei A. Kirilenko, Albert S. Kyle, Mehrdad Samadi, and Tuzun, Tugkan, The Flash Crash: the impact of high frequency trading on an electronic market, *Working Paper, 2014.*

Eric Budish, Peter Cramton, and John Shim, The high-frequency trading arms race: Frequent batch auctions as a market design response, *Working Paper,* 2013.

Andersen Torben and Oleg Bondarenko, VPIN and flash crash, Journal of Financial Marekts,2014.

Andersen Torben and Oleg Bondarenko, Reflecting on the VPIN dispute, *Research letter*, 2014.

CFTC Response to Allegations Pertaining to the Office of Chief Economist, February 2014.

Erick Aldrich, Joseph Grundfest, Gregory Laughlin, 2016, The Flash Crash: A New Reconstruction

Kerry Back, Kevin Crotty, and Tao Li, Estimating the Order-Flow Component of Security Returns, *Working Paper,* 2014

Andersen Torben, Oleg Bondarenko, Albert S. Kyle, and Anna Obizhaeva, Intraday trading invariance in the E-mini S&P 500 futures market, *Working paper*.

Angel, James, Lawrence Harris, Chester Spatt, Equity trading in the 21$^{st}$ century. 2010

Ananth Madhavan. VWAP Strategies. 2002.

OPTIONAL VIDEO:

## Questions for Write-Up. Answer starred questions (1 page):

1. *According to the regulators, does high-frequency trading destabilize financial markets and make them more volatile?

2.  *In your opinion, which of the markets, the U.S. E-mini futures market or the U.S. equity market, functioned better during the events of May 2010? What was the main reason?

3.  *The report regarding the market events of May 6, 2010 was the combination of the separate work done the staffs at the CFTC and the SEC. Which of them did a more professional analysis?

4.  According to Larry Harris, what is the most harmful in high-frequency trading and how to deal with that issue?

5.  Would high-frequency traders prefer to trade in markets with high or low tick size?

6.  How is high frequency trading in the futures markets different from high frequency trading in stock markets?

7.  Would Fischer Black like the idea of a VWAP order? Are VWAP orders a good idea? Can a dishonest agent exploit a naive customer who places a VWAP order?

# WEEK 3 (TUESDAY, NOVEMBER 13): MICROSTRUCTURE THEORIES

TOPICS: public versus private information; market efficiency; adverse selection; noise trading; insider trading; return volatility; market liquidity; market depth; inventory models; perfect competition; imperfect competition; smooth trading.

**REQUIRED READING:**

**\*Hasbrouck, Joel. "Securities Trading: Principles and Procedures." (2016). Part III, Chapters 9-12, pp. 67-110; chapter 16, pp. 135-140.**

**\* Fischer Black. "Noise".** ***The Journal of Finance*, 41(3):529–543, 1986.**

**\* Albert S. Kyle. "Continuous auctions and insider trading".** ***Econometrica*, 53(6):1315–1335, 1985.**

OPTIONAL READING:

Stephen F. LeRoy, Efficient capital markets and martingales, *The Journal of Economic Literature*, 27(4), 1583-1621, 1989.

John F. Muth. Rational expectations and the theory of price movements. *Econometrica*, 29(3): 315–335, 1961.

Albert S. Kyle. Informed speculation with imperfect competition. *The Review of Economic Studies*, 56(3): 317–355, 1989.

## Questions for Write-Up. Answer starred questions (1 page):

1. *What explanations do the required readings provide why there is so much of trading?

2. *What is the efficient market hypothesis? What is adverse selection?

3. *According to Fisher Black, what is a good measure of market efficiency?

4. *How does Pete Kyle explain where market depth comes from?

5. What happens to bid-ask spreads when the number of noise traders increases? When the fraction of informed traders increases? When the quality of informed traders' information increases?

6. How do you expect the volatility of oil prices (or other assets) to change when many speculators decide to enter the market where it trades?

7. Suppose a large trader breaks a block into small pieces and executes the pieces over time? Would this inject a predictable trend into prices such that it might be profitable to trade against this trend?

8. If you knew the otherwise undisclosed positions of profitable hedge funds, could you make money using this information? If so, how?

9. If you knew the undisclosed inventory of market makers, could you make money on the basis of this information? If so, how?

10. What is the difference between permanent price impact and temporary price impact? Is Kyle (1985) a model about permanent impact, temporary impact, or both?

11. What would you expect the relationship between transactions costs and alpha to be for a very successful hedge fund with modestly scalable trading strategy?

# WEEK 4 (TUESDAY, NOVEMBER 20): INVARIANCE

TOPICS: Invariance; implementation shortfall; dimensional analysis; leverage neutrality.

**REQUIRED READING:**

**\*Hasbrouck, Joel. "Securities Trading: Principles and Procedures." (2016). Part IV, Chapters 14-15, pp. 115-140.**

**\*Mark Kritzman, Kyle, Albert S., and Anna A. Obizhaeva, A practitioner's guide for market microstructure invariance, 2014 (explanation of invariance for practitioners)**

**\* Albert S. Kyle and Anna A. Obizhaeva, Dimensional analysis, leverage neutrality, and market microstructure invariance, *Working Paper.*, 2017 (explanation of invariance for physicists) Available at SSRN: https://ssrn.com/abstract=2785559**

**\* Perold, Andre F. Shortfall: Paper versus Reality. *Journal of Portfolio Management*, 14(3): 4-9, 1988.**

OPTIONAL READING:

Albert S. Kyle and Anna A. Obizhaeva, Market microstructure invariance: Empirical hypotheses, *Econometrica* 84(4), 1345-1404, 2016. (explanation of invariance for econometricians)

Albert S. Kyle and Anna A. Obizhaeva, Market microstructure invariance: A Dynamic Equilibrium Model, *Working Paper.* (explanation of invariance for economists)

## Questions for Write-Up. Answer starred questions (2 pages):

1. *Is business time the same as volatility time?

2. *Is a bet the same as as a trade, or transaction, or an order?

3. *According to invariance, if one compares transaction costs for liquid and illiquid securities, would the dollar costs of executing the average bet be the same across markets? Would the percentage costs of executing the average bet be the same?

4. *According to invariance, what is a good measure of liquidity?

5. *Calculate "implementation shortfall" for the following scenario: At 10:00 a.m., the stock price is $40.00 and you decide to buy 100,000 shares. During the next six hours, you purchase 80,000 shares at an average price of $40.50. At 4:00 p.m., the market closes at a price of $43.00 and you cancel the unexecuted balance of 20,000 shares.

6. *Calculate implementation shortfall for the following modified scenario: At 10:00 a.m., the stock price is $40.00 and you decide to buy 60,000 shares. You think you might buy some more, but you are not sure how much more. At noon, you have already bought 60,000

shares at an average price of $40.25; the current stock price is $41.00; you increase the size of your desired purchases for the day from 60,000 shares to 100,000 shares. Over the course of the rest of the day, you buy 20,000 additional shares at an average price of $41.25. At the 4:00 p.m. close, the price is $43.00, and you cancel the unexecuted balance.

7. *What is interesting about the answers to the two previous questions? What conceptual issues with implementation shortfall do the two previous questions address?

# WEEK 5 (TUESDAY, NOVEMBER 27): MARKET CRASHES.

TOPICS: Past and future market crashes; Ruble crisis of December 2014; financial stability, circuit breakers, price speed bumps.

**REQUIRED READING:**

**\*Kyle, Albert S. and Anna A. Obizhaeva, "Large bets and stock market crashes".** *Working paper*. **2014.**

**\* Kyle, Albert S. and Lee, Jeongmin, "Toward a Fully Continuous Exchange" (July 4, 2017). Available at SSRN:https://ssrn.com/abstract=2924640**

OPTIONAL READING:

Anna Obizhaeva, A little note on Chinese market crash of 2015

Tugkan Tuzun, "Are leveraged and inverse ETFs the new portfolio insurers?", *Working paper*. 2014.

Jiangze Bian, Zhiguo He, Kelly Shue, Hao Zhou, Leverage-induced fire sales and stock market crashes, 2018

Kyle, A.S., Obizhaeva, A.A. and Wang, Y., 2017. "Smooth Trading with Overconfidence and Market Power." *Review of Economic Studies*, forthcoming

OPTIONAL VIDEO: Pete Kyle on crashes. https://www.youtube.com/watch?v=pnsoQOuP_Vs

## Questions for Write-Up. Answer starred questions (2 pages):

1. *According to Kyle and Obizhaeva, what are the two types of market crashes and which of them is more difficult to deal with?

2. * In which situation the price impact will be larger: when 1% of average daily volume (ADV) is dumped into the market of E-mini S&P500 futures or into the market of some illiquid stock?

# WEEK 6 (TUESDAY, DECEMBER 4):  FICC MARKETS.

TOPICS:   Market microstructure of fixed income markets; Flash Rally of Oct 15, 2014; trading liquidity and funding liquidity; economics behind fixing, LIBOR and WMR manipulations.


**REQUIRED READING:**

**\* Anna Obizhaeva, "The Russian ruble crisis of December 2014", Voprosi Ekonomiki, 2016.**

**\*CFTC, SEC, Treasury, Joint Staff Report on the U.S. Treasury Market on October 15, 2014**

**\* Larry Harris, Albert S. Kyle, Erik Sirri, FAJ, 2015, The structure of trading in bonds market.**


OPTIONAL READING:

Bank of England, Fair and Effective Markets Review,  2015

Libor Scandal,. Review of Banking and Financial Law. 2012

Larry Harris, "Transaction Costs, Trade Throughs, and Riskless Principal Trading in Corporate Bond Markets," October 22, 2015. http://dx.doi.org/10.2139/ssrn.2661801

Геннадий Пифтанкин, 2015, "Микроструктура российского валютного рынка и валютный кризис в декабре 2014 года"


OPTIONAL VIDEO:
Roundtable on banking and financial markets, July 13, 2015:
www.c-span.org/video/?327066-1/roundtable-banking-financial-markets

## Questions for Write-Up. Answer starred questions (1 page):

1. \*What are FICC markets?

2. \*Based on Bank of England's Review, what are common themes in recent FICC misconduct cases?

3. \*Compare the report on the flash crash in U.S. Treasuries market on October 15, 2014 with the report on the flash crash on May 6, 2010. Which of the two reports is better? Which recommendations are stronger?

4. What evidence does the report present to back up its assertion that the U.S. Treasury market is the "deepest and most liquid government securities market in the world," e.g., more liquid than the government securities markets of the U.K., China, Japan, or Germany?

5. Why do the Report writers believe that a circuit breaker which triggers during extremely rapid price movements mitigate activity or price movements which may not accurately reflect fundamental forces of supply and demand?

# WEEK 7 (TUESDAY, DECEMBER 11):        INSIDER TADING AND PRICE MANIPULATION. REVIEW.

TOPICS:   Insider trading, price manipulation, spoofing, layering, Sarao's case; review.

REQUIRED READING:

**\*Hasbrouck, Joel. "Securities Trading: Principles and Procedures." (2016). Part IV, Chapters 10 and 12, pp. 79—89 and 102-110, Chapter 17, pp. 144-155.**

**CFTC complaint, Sarao case, April 17 2015, with Appendix**

OPTIONAL READING:

Andrew Verstein, Revolution in manipulation law: The new CFTC rules and the urgent need for economic and empirical analysis, 2013.

Albert S. Kyle and S. Viswanathan, How to define illegal price manipulation? *American Economic Review Papers and Proceedings*, 98 (2), 274-279, 2008.

CFTC Complaint, Igor Oystacher case, 19 October 2015

Mikle Coscia case

## Questions for Discussion. Answer starred questions (1 page):

1.  Did the CFTC think that Sarao created a Flash Crash? List CFTC arguments.

2.  Would Sarao agree with CFTC charges? List his arguments.

## *PROJECT 1: NASDAQ ``Odd Eighths'' Controversy in 1994*

### *(Deadline: 21 Nov, 23:00)*

The data are ZIP files containing compressed CSV files of tick data for specific assets. Download the file or files into your own computer. Keep track of the directory into which you unzip the files. (*Note:* You may want to look at the data in the CSV files. If you use Microsoft Excel, the time stamps may be automatically converted into an incorrect format. If so, the try Notepad or some word processing software to view relatively large CSV files as unformatted text).

This homework relates to the discussion we had in class on the papers of Christie and Schultz (1994) and Christie, Harris, and Schultz (1994) papers. You are asked to do the following:

### 1. Data Cleaning

Data sets often have errors. Data errors in market microstructure can be classified into various types:
> • Data can be logically invalid, such as the date April 31, 2018.
> • Market microstructure data can violate reporting or trading rules or conventions, such as quotes being in "eighths" of a dollar (increments of $0.125) or in one-cent increments. For example, a quote of $31.125 (Thirty-one-and-one-eighth) was allowed in 1994 but is not allowed 2018. A quote of $31.17 was not allowed in 1994 but is allowed in 2018.
> • Data can also be valid but simply incorrect. For example, a trader might erroneously report of trade price of $999.50 when the trade occurred at $99.50 and a report of $99.50 was intended.

Data errors can be intrinsically ambiguous, even after they have been reported and possibly corrected.
> • For example, a customer may think he sells 5000 shares of stock at $99.50, but the dealer thinks the sale was for 500 shares at $99.50. The dealer reports a sale of 500 shares. The customer and dealer get into an argument, which is resolved weeks later, perhaps in favor of the customer.
> • The price of a share may collapse, either due to bad news or to a pre-announced stock split. A trade may occur at a very low price because a market order (or stop loss order) triggered a sale when there are no competitive bids in the limit order book.

It is good practice in data analysis to clean data with a self-documenting program which takes raw data as an input and cleans the data with transparent code. This makes it possible to replicate what you have done. In some cases, you may want to delete bad data; in other cases, it is better simply to label bad data as such to exclude it from the analysis. You are asked to find data which is potentially erroneous, even if the potential error is ambiguous.

Obtain the trade and quotes files containing data on 30 stocks for the year 1994. Of the 30 stocks, there are 10 from Nasdaq, 10 from NYSE, and 10 from Amex. It may in a ZIP file.

• Perform various checks for invalid data. Label potentially invalid data as potentially invalid.

• Perform other checks to identify logically valid but potentially erroneous data. These include prices or quantities which are large outliers. Label identified data as outliers.
• Regular trading hours are from 9:30am(open) to 4:00pmEST (close). Label for removal quotes that occur before the opening of trading and after the close.
• You may only want to use NASDAQ trades and quotes when examining NASDAQ stocks. Make sure you understand exchange codes. You can find the relevant exchange codes at http://www.nyxdata.com/doc/2552.
• Exclude outliers, such as rows with missing information, or quotes or sizes with a value of zero.
• Adjust quote sizes to the number of shares. In the file, the sizes are given in hundreds of shares.
• You will want to exclude abnormal quote modes. You can find the list of modes in the same document as the exchange codes. Briefly document what modes you decide to exclude and why. (Note: It is good practice in data analysis to clean data with a self documenting program which takes raw data as an input and cleans the data with transparent code. This makes it possible to replicate what you have done.)

## 2 Analysis

You are asked to determine whether dealer quoting behavior changed during the year 1994. Perform the following analyses after excluding invalid data, outliers, quotes outside of regular hours, non-NASDAQ quotes (if any), and abnormal quote modes.

• Plot the average observation-weighted bid-ask spread and average time-weighted bid-ask spread for all stocks for each trading day in the year 1994. (You may use smaller increments of time, such as 5-minute intervals instead of daily intervals.)
• Determine whether the average bid-ask spreads before and after the structural break are statistically different. Can you justify the structural break with a formal statistical test or with informal probabilistic logic? (Note: The idea here is to determine whether the market might have changed its behavior by chance.)
• Describe changes that occurred in the average quoted quantities both before and after the structural break.
• Determine for what percentage of the time the bid-ask spread was 1/8, 1/4, 3/8, or a different amount, both before and after the break. Illustrate graphically.
• For extra credit, provide any other analysis you think may be useful about the behavior of bid and ask prices during this time period.
• For extra credit, examine whether trade size and distribution of trades across even and odd eighths changed during the year 1994.
• For extra credit, do anything else you think may be interesting.

The write-up cannot simply represent a Python code. It has to be done in a manner consistent with good business writing style, starting with a short executive summary and ending with conclusions.

## PROJECT 2: Flash Crash in 2010

### (Deadline: 27 Nov, 23:00)

This homework relates to the discussion about Flash crash. The data are ZIP files containing compressed CSV files of tick data for E-mini futures and ETFs. Questions are meant to be open-ended. Good answers might vary a great deal from student to student.

1. Download the ZIP file. Unzip the ZIP file, which contains four CSV files of quotes and trades for the SPDR ETF and the S&P 500 E-mini futures contract.

2. Clean the data:
   - Excludes bad observations with price or size of 0.
   - Excludes "crossed market" and "locked market" quotes.
   - Address the time zone issues. Note that the timestamps of the SPY data are New York time, but the timestamps of E-mini data are UTC time.
   - Excludes observations out of trading hours, which are 9:30–16:00 in New York time.
   - Examines duplicates in each data set.
   - Deals with NAs in the futures data (because one row of futures data only reports
   - the update of bid or ask).

3. Initial analysis:
   - Plots trading volume minute by minute.
   - Plots realized volatility the E-mini and SPY over various time periods and for various exchanges.

4. Normalize the futures and spot prices to 100 by dividing each price by the first price of the day, then multiplying by 100. Plot the normalized futures and SPDR prices on the same plot for the entire day. Plot the future and SPDR prices for the time period of the flash crash. Did the arbitrage relationship between the E-mini futures contract and the SPDR ETF break down during the flash crash? If so, did the breakdown occur at the beginning, middle, or end of the flash crash?

5. Plot bid-ask spreads and market depth for both the E-mini and the SPDR. How do bid-ask spreads and market depth change during the day?

6. Plot average trade size for E-mini and SPDR during the day and the flash crash. Did average trade size change during the Flash Crash?

7. If typical volatility is one percent per day, how much greater than typical was volatility on the day of the flash crash? Why does realized volatility vary depending on the exchange used to construct the graphs of cumulative volatility in the sample program?

8. Do price changes on some exchanges lead price changes on others?

9. Examine the correlation of price changes for E-mini and SPDR for different horizons. Do you observe any patterns? How can you explain them?

10. For extra credit, do anything else you think may be interesting.

*PROJECT 3: Invariance*

*(Deadline: 11 Dec, 23:00)*

This homework is about invariance. Download the compressed CSV files for trades and quotes from dropbox. One file contains all trades for all tickers for one day. The other files contain quotes for all tickers beginning with the letter "A". There are several quote files because large files might be difficult to read on computers with 4–8 gb of memory.

1. Download data for tickers with the letter "A". Read each of the several quote files one by one, perform checks on the data for consistency, remove or disregard quotes before or after regular trading hours, remove or disregard obvious examples of bad data, and keep only regular quotes (quotes where (qu_cond == 'R').

2. For each ticker symbol, calculate a table of aggregate statistics including number of trades, sum of dollar spreads, sum of bid share size, sum of ask share size, and any other statistics which seem interesting. As you read through each file in a loop, save the aggregate statistics but remove the previously read large data table of raw data to save computer memory.

3. After extracting a data table of aggregate statistics for each of the quote files, merge the data tables together into one data table. Then merge the summed aggregate statistics together by ticker symbol. This is necessary because, even though the raw files are sorted by ticker symbol, data on some ticker symbols appears at the end of one file and at the beginning of the next. Next, convert the aggregate sums into averages by dividing by the number of observations (quotes). Explain why averages are calculated as a last step in this manner rather than calculating averages (mean) when the data were first processed.

4. Perform the above steps separately for each exchange, creating for each exchange one merged dataset with one line of aggregate and mean statistics for each exchange.

5. Merge these data tables with a data table containing aggregate dollar volume and realized volatility (both obtained from the trade data).

6. Think about how to test invariance cross-sectional predictions for dollar spreads. For each exchange and for all the data combined, draw corresponding plots with dollar spreads on the vertical axis. Save all of the plots into a PDF file. There should be files for each exchange and files for all exchanges combined. Include with the plot a fitted regression line with a slope estimate. What does market microstructure invariance predict that the slope coefficient should be (approximately)? Explain what approximations or shortcuts are being taken using this approach. Discuss anything interesting you can spot on the figures.

7. Think about how to test invariance results for share quote sizes. For each exchange and for all the data combined, draw corresponding plots with share quote sizes on the vertical axis. Save all of the plots into a PDF file. There should be files for each exchange and files for all exchanges combined. Include with the plot a fitted regression line with a slope estimate. What does market microstructure invariance predict that the slope coefficient should be (approximately)? Discuss anything interesting you can spot on the figures.

8.  For extra credit, discuss the problem and suggest potential approaches for testing invariance predictions in fragmented markets where each security may be traded at multiple exchanges with different spreads and quote sizes.

9.  For extra credit, do anything else you think may be interesting.