Trading Equities: Market Structures and Costs

Bernt Arne Ødegaard

Dec 2016

1 Introduction

Topic of this lecture: "Market microstructure for investors." The *trading* of financial assets, how has it changed, and what do investors need to know about the topic.

1.1 Overview of the lecture notes

- How market places have changed
- How methods of trading have changed
- Measuring the cost of trading - Some estimates
- Regulation
- Lessons for trading
- Gazing into the Crystal Ball
- Further Reading

2 How the market places have changed

In this section we show two snapshots of how equity markets are organized, around year 2000, and now (ca 2016). What are the key features of trading at those two points of time? What has changed?

2.1 Year 2000

USA
Two major stock exchanges

- NYSE, the largest, blue chip stocks. Trading done physically on the floor, with electronic support (limit order books).
- Nasdaq, smaller companies, tech stocks, etc. Electronic limit order book, human interaction.
- Minimum spread: sixteenths (down from eights in 1997)

The stocks listed on NYSE/Nasdaq also trade on

- ATS – Alternative Trading Systems (alternative limit order books at major investment houses)
- ECN - Crossing Networks (between-investor trading using prices from NYSE/Nasdaq)

However, limited in scope.

Europe

- The national exchanges (LSE, Paris, Deutsche Börs....) dominant.

Norway

- Oslo Stock Exchange dominant.

2.2 Now (ca 2016)

USA

- Regulation (SEC): Reg NMS(2006): Monopolies disappeared. Trading can happen at lots of registered market places.
  - Important for traders: NBBO (National Best Bid and Offer): Baseline Price. Obligation to trade at best prices.
- Almost all trading electronic.
  - Big markets (NYSE, Nasdaq, BATS, etc): Limit order markets
  - Numerous alternative market places (Dark Pools). Not necessarily independent price discovery
- All trading is fragmented.
- Uniform spread: One cent

Europe

- Regulation (EU): MiFiD: Mandate Competition between market places.
- Result: Fragmentation of trading
  - Major markets: Electronic limit order markets, little manual trading
  - Sattelite markets (Dark Pools)
- No Europe-wide reporting of single price (ie. no NBBO)
- Spread schedules with ticks below US one cent minimum tick (spread)

Three types of trading markets.

1. Regulated markets (RMs)
2. Systematic Internalizers (SIs)
3. Multilateral trading facilities (MTFs)

Lead to entry of new MTFs

- Chi-X (Mar 2007)
- Turquoise (Mar 2007)
- BATS Europe (April 2008)

- Nasdaq OMX (Sep 2008)
- NYSEArca (March 2009)
- Burgundy (May 2009)

Existing exchanges merge.
Euronext, merger of Paris, Amsterdam, Brussel and Lisbon
Stockholms OMX Scandinavian and others
LSE, aquires Borsa Italiania, Turquoise
Additionally, numerous Dark Pools, other OTC markets.
Difference with the US: No notion of market-wide best bid and offer, as in the US markets.

Norway

- Regulation (EU): MiFiD: Applies also to Norway (even if not in EU).
- Result: Fragmentation of trading
  - Oslo Stock Exchange currently about a third of total trading in Norwegian Stocks.
  - Other market places for Norwegian Shares
    - Other Exchanges (Stockholm, BATS, Chi-X...)
    - Satellite markets (Dark Pools/OTC trading)
- Spread schedules with ticks below US one cent.

2.3 Some Examples

USA Let us look at some numbers for fragmentation of trading. Figure 1: NYSE and figure 2 Nasdaq both show distribution of trading at two points in time, 2005 and 2014, for respectively NYSE and Nasdaq stocks.

Figure 1 Fragmentation of trading of NYSE stocks

Market Share – NYSE Stocks

February 2005		February 2014	
NYSE	78.9%	NYSE	20.1%
NASDAQ SuperMontage	2.1%	NASDAQ	14.2%
CHX	2.1%	NYSE Arca	8.4%
PCX Arca	1.7%	BATS Z	7.0%
BSE (now Nasdaq BX)	1.1%	EDGX	6.6%
NSX	0.7%	NASDAQ BX	2.9%
Phlx (now Nasdaq PSX)	0.4%	EDGA	2.5%
Dark ATS		BATS Y	1.9%
and Broker-Dealers	13%	ADF	0.5%
		NASDAQ PSX	0.4%
		CBSX(no longer operating)	0.4%
		CHX	0.3%
		NSX (no longer operating)	0.2%
		Dark ATS	
		and Broker-Dealers	34.6%

Source: SEC Memorandum to Market Structure Advisory Committee on Rule 611 of Regulation NMS. 2015

Figure 2 Fragmentation of trading of Nasdaq stocks

Market Share – NASDAQ Stocks

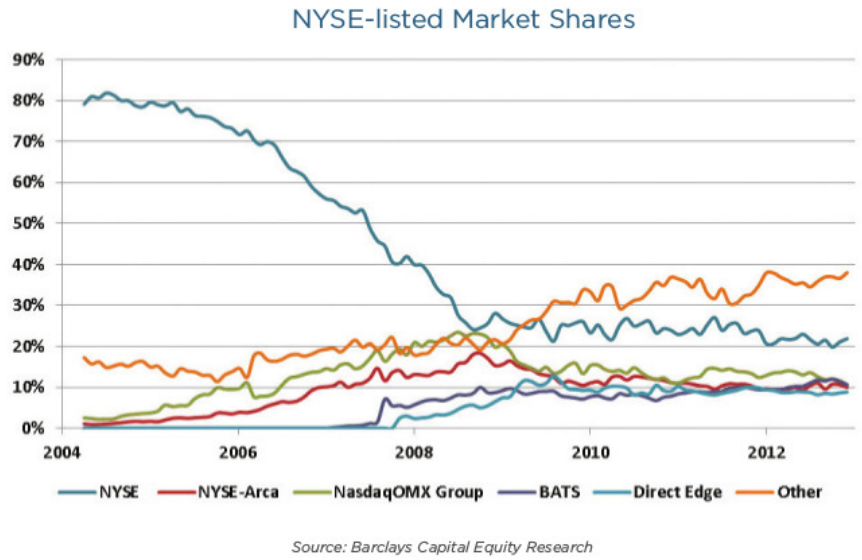
February 2005		February 2014	
NSX (reporting Inet ECN trades)	25.3%	NASDAQ	26.2%
Nasdaq SuperMontage(estimated)	19.2%	NYSE Arca	9.0%
PCX Arca (now NYSE Arca)	19.0%	EDGX	8.7%
Brut ECN(estimated)	6.0%	BATS Z	8.0%
CHX	0.5%	NASDAQ BX	2.7%
NASD ADF	0.5%	EDGA	2.6%
Amex (now NYSE MKT)	0.1%	BATS Y	1.8%
Dark ATS		FINRA ADF	0.7%
and Broker-Dealers	29.4%	NASDAQ PSX	0.5%
		CBSX(no longer operating)	0.4%
		NYSE MKT	0.3%
		NSX(no longer operating)	0.3%
		CHX	0.3%
		Dark ATS	
		and Broker-Dealers	38.6%

Source: SEC Memorandum to Market Structure Advisory Committee on Rule 611 of Regulation NMS. 2015

Let us supplement these numbers with the time series of market shares, taken from Angel, Harris, and Spatt (2015). The first, figure 3 shows how the market share of NYSE of trading in NYSE listed firms have fallen, with other, new market entrants taking up the slack. Both BATS and DirectEdge have managed to get large

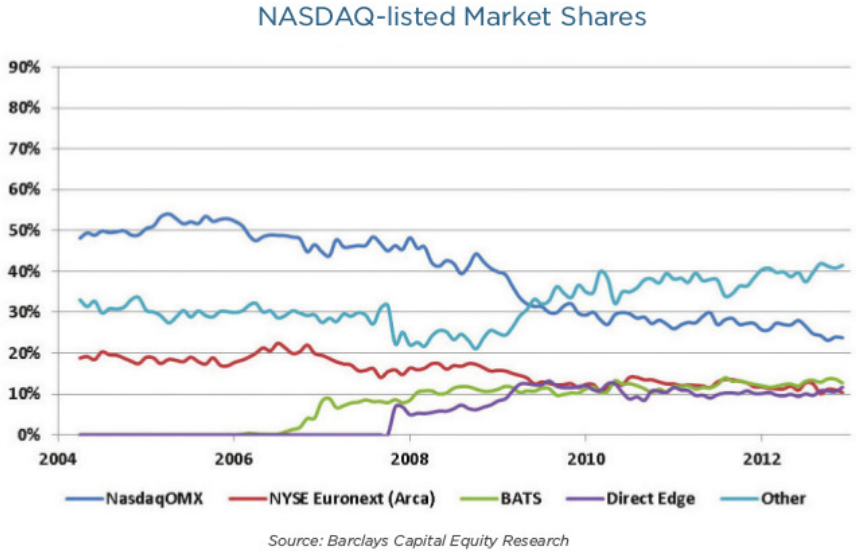
market shares over short times. This is even truer for Nasdaq shares, as shown in figure 4. The total share of trading happening in dark pools is shown in figure 5, and it seems to have levelled off at slightly less than 15%.

Figure 3 Market Shares - NYSE listed firms



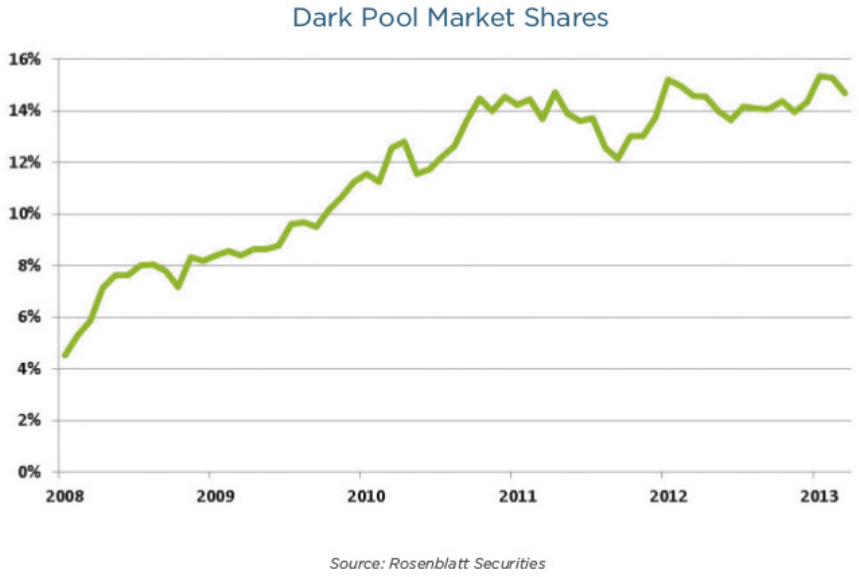
Source: Angel et al. (2015).

Figure 4 Market Shares - Nasdaq listed firms



Source: Angel et al. (2015).

Figure 5 Dark Pool Market Shares



Source: Angel et al. (2015).

Europe In Europe the situation is similar. Figure 6 shows examples for the largest stocks on the French exchange (the CAC40) and the London Stock Exchange (FTSE100). The table only show trading on listed exchanges with their own limit order books (price discovery). It does not include the type of Dark Trading shown for the US. According to Foucault, Pagano, and Roell (2013), for the European markets, 11% of trading (in addition to the numbers shown in figure 6 are traded OTC.

Figure 6 Fragmentation of European equity trading, 2011

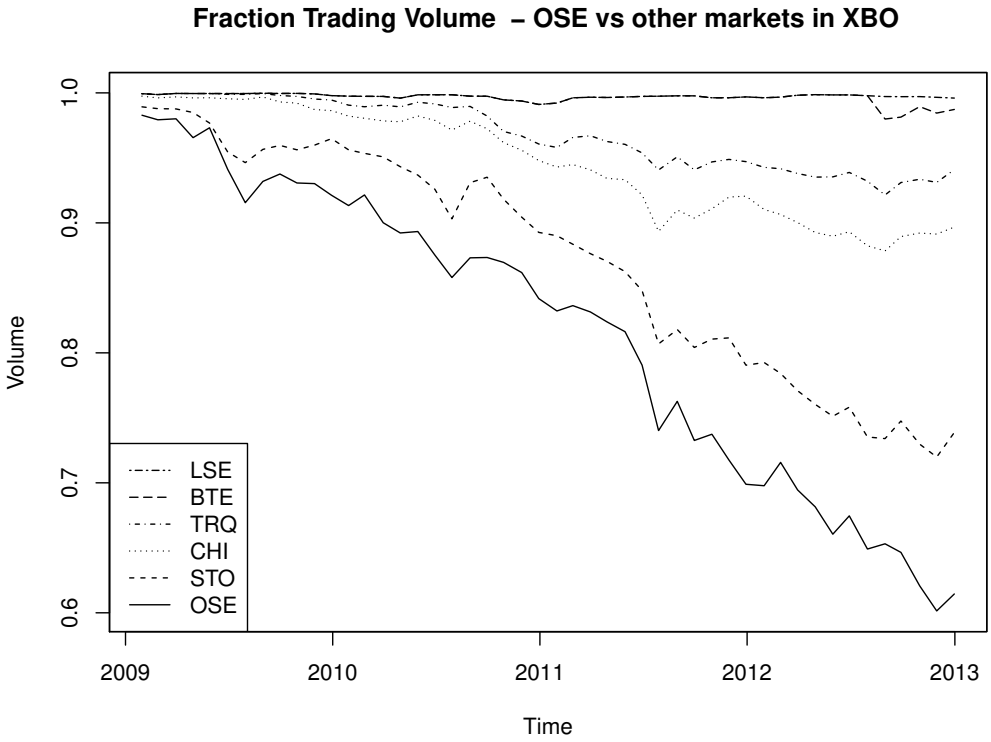
Trading Platform	Market Share	
	CAC40 Stocks	FTSE100 Stocks
Euronext	60.5%	n.a.
LSE	n.e.	54.77%
Chi-X	24.95%	30.43%
Turquoise	5.67%	7.93%
BATS Europe	4.31%	6.66%

Market Fragmentation in Europe, CAC40 (French) stocks and FTSE100 (UK) stocks.
Source: Foucault et al. (2013)

Norway Let us also look at how trading of *Norwegian* stocks has fragmented in recent years. We use data from Thomson Reuters to give some evidence on the extent of fragmentation. Thomson Reuters provides an Europe-wide summary of trading through their "XBO." This is a summary of the trades of the same stock on public limit order markets. In figure 7 we use these data to illustrate the evolution of fragmentation of trading in stocks with a main listing at the OSE. The figure show the proportions of the total trading volume done at the various market places. We see that the proportion of trading at the OSE has been falling throughout the period. Note that this does not include trading on OTC market places.

Figure 7 Norwegian stocks traded at the OSE and other lit markets

The figures illustrate where Norwegian stocks are trading. The figure shows fractions of trading in each market. The bottom line is the fraction of the XBO traded at the OSE. The other exchanges adds to the total. The fractions sum to one (we have left out some very small market places). Here OSE is the Oslo Stock Exchange, STO is Nasdaq OMX, BTE is Bats, TRQ is Turquoise, CHI is Chi-X and LSE is the London Stock Exchange.



Source Jørgensen, Skjeltorp, and Ødegaard (2017)

3 How methods of trading have changed

In 2000, the default method of trading involved human interaction, either on a floor (NYSE), or by human entering of traders into the limit order book. By now (ca 2016) much of the human interaction has been replaced with computerized trading.

3.1 High Frequency Trading (HFT)

Terms of usage:

- **Algoritmhic Trading**: Trading where computers involved in trading *decisions* (not just as an order entry system.).
- **High Frequency Trading** (HFT): A subset of algorithmic trading where *speed* (of communications, of trading decisions) is important determinant of profits from trading strategy.

Informally, High Frequency Trading (HFT) is used more broadly to refer to all computerized trading.
Driving force behind the computerization: Getting rid of humans in the loop lowers trading costs substantially. Replacing (highly paid) brokers with computers.
Computers win because:

- Computers have a quick and wide attention span: They can react faster to new information.
- Computers are disciplined, they do only what they are programmed to do. (Which may also be a liability).
- Computers do not forget.
- Computers do not argue about office placement, holidays, pay, $\cdots$

So, finance is the last industry to enter the industrial revolution.
HFT has become a very controversial topic, with lots of negative press.
Will give a more balanced view.
Everybody agrees:

- HFT opens for more efficiently performing the traditional role of financial intermediaries:

  Market Making – Providing liquidity, and being compensated for it (earning the spread).

Also agreement about:

- HFT gives advantages to the fastest traders.

What is disagreement about:

- Are the advantages to the fastest trades of a nature that we should worry? (and regulate)?

Some of the arguments in this debate:

- "The speed advantage removes the level playing field."

  The playing field of market trading has never been level. The closer you are to the market the more informed you are. The old time Specialist on the NYSE had a lot of information about trading interest that nobody else knew.

  Here we need to know what in particular is different about HFT.

- "HFT traders frontrun other traders"

  The main theme of Michael Lewis (2014)' book *Flash Boys*.

  However: Frontrunning in its pure form (trading ahead of your customers orders) is illegal.

  Rather, talking about

  - HFT traders uncovering direction of trader interest.
  - HFT traders trading in correlated assets (Buys in Apple likely to be followed by Buys in Microsoft, IBM, etc?)

- "Some HFT trading practices is market manipulation"

  Market manipulation is illegal. However, very hard to "prove intent"

  Hypothesized games HFT traders play.

  - **Smoking**

    Smoking involves posting short-lived quotes inside the spread to tease other (slow) traders to enter.

    Suppose a HFT trader wants to sell. Will then post short-lived quotes *inside the spread* that are not meant to be executed to tease buyers into crossing the spread

    Let us look at an illustration, shown in figure 8

    Note however that this is not risk free from the HFT point of view, there may be hidden orders in the book that is hit by the quotes.

  - **Spoofing**

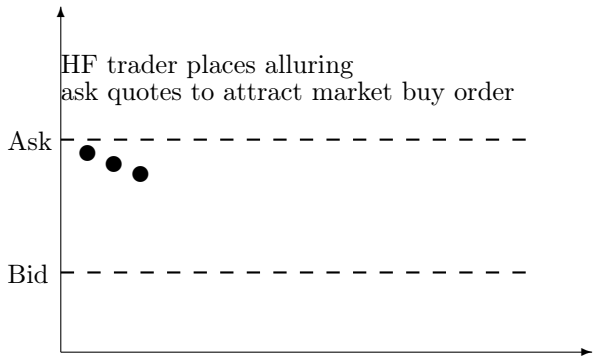
    Spoofing involves posting quotes *away* from the spread to give an illusion of lots of trading interest.

    Suppose the HFT really want to buy. Will then place lots of sell orders. See figure 9

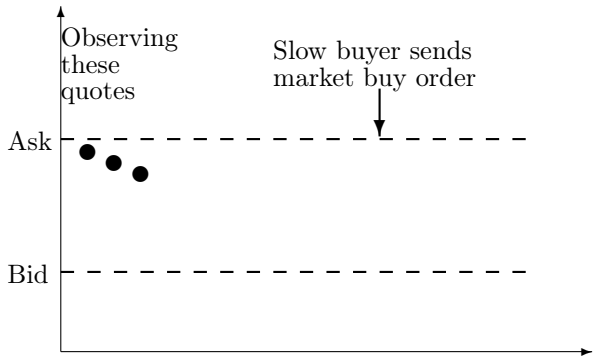
    Also not risk free, suppose a large order takes out the best ask, may end up selling instead of buying...

Figure 8 Smoking

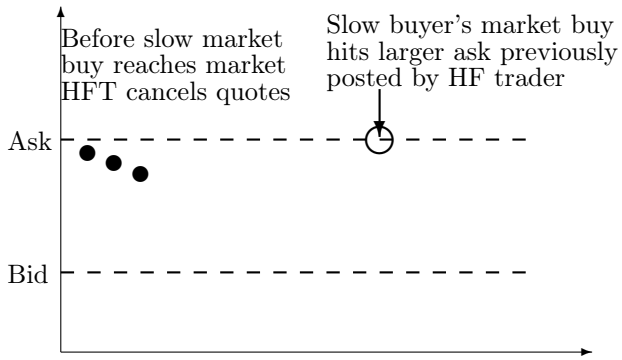
Step 1



Step 2



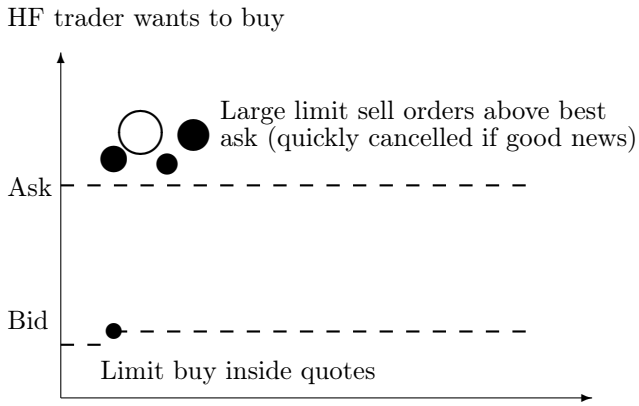
Step 3



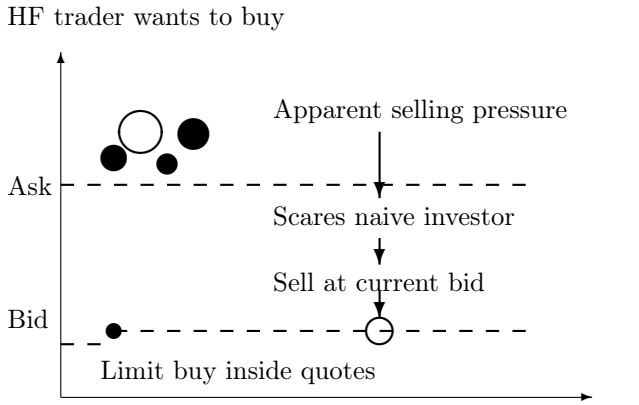
Illustrating the concept of *smoking*. Pictures borrowed from Biais (2011).

Figure 9 Spoofing

Step 1



Step 2



Illustrating the concept of *spoofing*. Pictures borrowed from Biais (2011).

- "Automated trading leads to a fragile system" (or – Crash testing market structures)

  The Flash Crash was an event where trading in New York and Chicago broke down for half an hour on May 6, 2010. The Dow Jones index lost 9% within minutes, but recovered after 38 minutes.

  Critics point to the flash crash as an example of the systemic fragility of an automated market.

  "Systemic Risk"

- Is there a conclusion about HFT being good or bad?

  Not really.

  Primary evidence: Substantial lowering of measured transaction costs as HFT takes over. (Will see later)

  But can not rule out that some of the negatives are large enough in magnitude to offset the lower transaction costs.

  This has lead to the two terms:

  – HFM – High Frequency Market Making – Benign HFT.
  – HFB – High Frequency Bandits – e.g. manipulative practices.

High Frequency Trader quits Some recent evidence: High Frequency Traders are getting out of the HFT business. See example from Financial Times in figure 10

Figure 10 High Frequency Traders Quit Business

Top high-frequency trader Teza to quit proprietary trading
Group founded by Misha Malyshev switches focus to quantitative hedge fund

One of the biggest and fastest traders in financial markets is abandoning its core business after its revenue engine stalled, a sign of the challenge of adapting in markets that unfurl in nanoseconds.

Teza Technologies of Chicago plans to exit its proprietary trading business in the next six months to focus on building up a quantitative hedge fund that manages more than $1bn, company executives said.

The new tack comes after net revenues at the proprietary business, which bets Teza's own money in markets from futures to bonds, steadily declined from about $250m four years ago to $80m in 2015, according to three people familiar with the figures. In 2016, the business has struggled to make a profit, the people added.

''Generally, it is harder to make money,'' Misha Malyshev, Teza chief executive, said in a rare interview.

From the Financial Times November 11, 2016.

3.1.1 Consequences of HFT: Colocation

Let us look at some consequences of HFT, to show what has changed. Colocation involves getting as close as possible to the exchange's computers. What has happened to the NYSE?

Colocation We are used to thinking about the NYSE like figure 11, a beehive of activity (this is in 1987).

Figure 11 New York Stock Exchange, floor in 1987

Photo Source: WSJ.

However, after the computers are taking over, the floor of the NYSE looks like figure 12.

Figure 12 New York Stock Exchange, floor in 2008

Photo Source: Wikipedia.

Instead, the activity has moved across the river, to Mahwah, New Jersey, where NYSE has built its data center, as shown in figures 13 to 15.

The NYSE has built a revenue stream from offering traders direct access to the exchange's computers, from computers in the same building. Figure 16 illustrates how NYSE are selling their colocation services.

The advent of colocation also illustrates how *exchanges* are chaning their business model.

Figure 13 New York Stock Exchange, Move to Mahwah

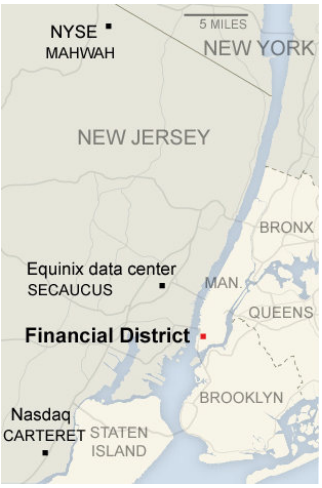


Figure 14 New York Stock Exchange, Mahwah, NJ. Entrance

Photo Source: NYSE

Figure 15 New York Stock Exchange, Servers, 2016

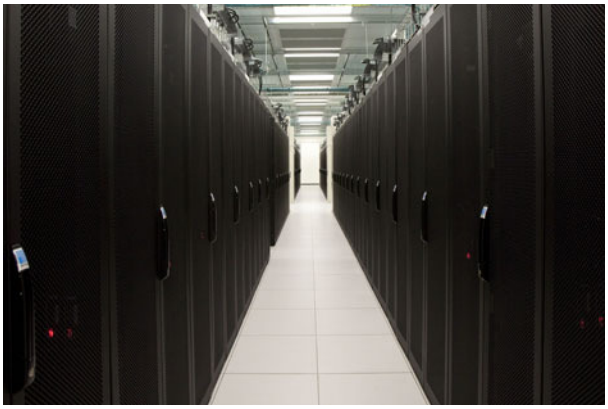
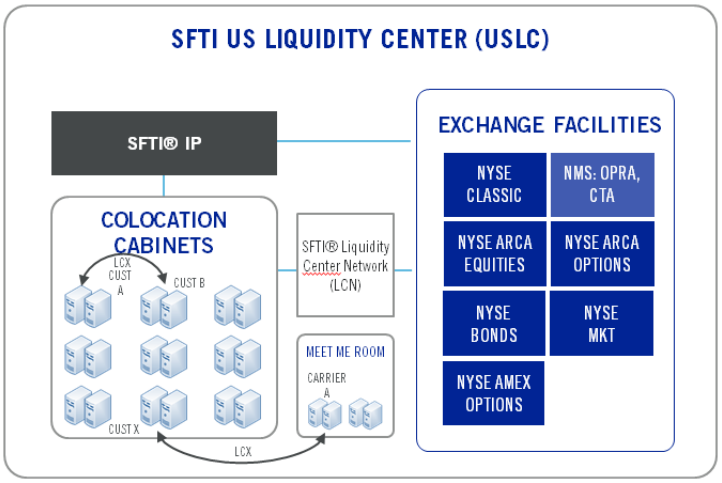


Photo Source: NYSE.

Figure 16 New York Stock Exchange Offer colocation to customers



Source: NYSE.

3.1.2 Consequences of HFT: The Race for Speed

Transmitting Information – Microwaves Let us show an illustration of the search for speed of moving information from one place to the next for the HFT traders.

Chicago to New York Any reader of "Flash Boys" will remember the stories of driving through Pennsylvania searching for the straightest line to lay optic fibre from Chicago to New York.

What is probably less known is that that technology lost out in very short order, to microwave transmitters, which, although they are low bandwith, are line-of-sight, and can therefore send information in a straighter line than fibre-optic cable, and therefore faster.

Laughlin, Aguirre, and Grundfest (2014) has an interesting picture that shows how the financial markets realization that microwave transmissions could improve speed lead to lots of spending on infrastructure. The picture on the left show the localization of applications for new microwave towers in the years before 2012. The picture on the right show the number of applications the next year, when it was realized that microwaves could be used as transmitters of financial information. The Chicago – New York corridor is prominent in the righthand picture, showing how much of the new bandwidth was concentrated here.

Figure 17 Licences for Microwave Towers

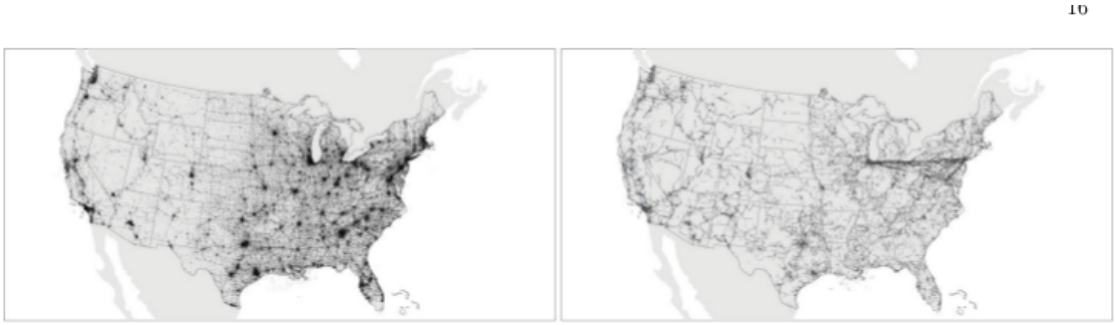


FIG. 10: A map of 93,600 paths present in 187,338 microwave site-based license applications for the two years prior to September 1, 2012 (left) and of 191,290 towers culled from the FCC database (right). Colors indicate frequency bands: 6 GHz (black), 11 GHz (charcoal), 18 GHz (dark gray), and other (slate gray).

Figure taken from: Laughlin et al. (2014)

The Richborough Mast, London and the continent The US is not the only place with scope for high-speed communications. In Europe trading is decentralized among various financial centers, two of the most important are Frankfurt and London.

If you visit the homepage of Vigilant Global, you will find the following proposal for the "Richborough Mast":

> *"We are seeking planning approval from Dover District Council to establish a new 322m tall communications mast on land adjacent to Richborough Energy Park, and provide a much-needed high speed data connection between financial institutions in the UK and Europe.*
>
> *The site was chosen as it is adjacent to an industrial site; is geographically remote and has a history of tall structures due to the areas previous use as a power station. It is also ideally located to achieve line of sight communications with existing infrastructure in Europe."*

together with documentation for why a 322 m mast is needed at the southeast coast of Britain, on the straight line between London and Frankfurt.

Figure 18 Connecting London and Frankfurt

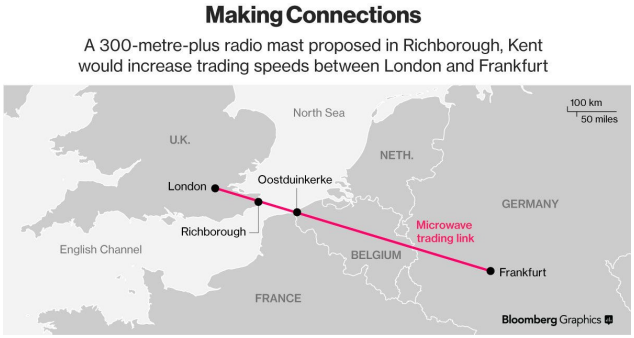


Figure source: Bloomberg, article 28 jul 2016.

Figure 19 Why a 322 m mast is needed to reach Europe

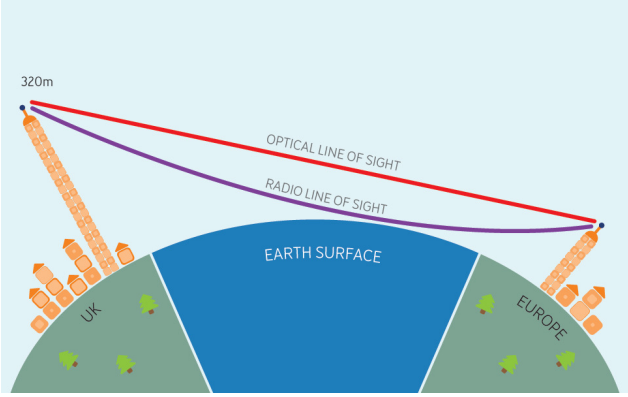


Figure from: Promotional material by Valiant.

In fact, there is now (Fall of 2016) actually two competing proposals, by two companies wanting to build a mast in this area. The original, by Valiant, is the "Richborough Mast." The second, by "New Line Networks," is the "Kings End Farm Mast".

3.2 How exchanges' methods of making money has changed

Historically, there was two major sources of revenues for exchanges (remember, exchanges used to be run like clubs by its members (traders)).

- Listing fee (from issuers)
- Fee per trade.

That has changed. Exchanges are now for profit companies, with more sources of revenue.

- Listing fee (still)
- Fee structure much more creative (and important for traders to be aware of)

  Examples:

  - Maker / Taker pricing
  - Payment for Order Flow
- Colocation and similar services, paying more gives better access to the exchange.
- Selling (Fast) Market Data another source of revenue.

Let us look a bit closer at two of these.

3.2.1 Maker Taker Fees

Maker/Taker Fees The make/take fee construction is a way for stock exchanges to compete for order flow from investors.

Exchanges charge liquiditity-demanding orders (marketable orders – demanding immediated execution) a fee (the "take fee"). This fee is offset by a rebate (the "make" rebate) for liquidity-supplying orders (nonmarketable limit orders). Exchanges make a significant part of their revenues from the difference between the fee and the rebate.

A motivation from the exchange's point of view is that the exchange wants the deepest market possible. By encouraging resting limit ordes (through the make rebate), the exchange encourage traders to choose their limit order book to place their resting orders.

Exchanges show a wide variety in their fee schedules, as shown in table 1, which shows the various fee schedules for US exchanges in November of 2012. Note that some exchanges actually have a negative fee, i.e. the *taker* is paying the fee.

The numbers in table 1 are indicative of the fees on each exchange, but some exchanges practice more differentiation by order type. The NYSE, for example, has a highly complex fee structure, shown in figure 2.

The practice of a maker/taker fee schedule has come under attack. For example, Angel, Harris, and Spatt (2011), in their survey of the state of the equity markets conclude:

> "Electronic trading raises some concerns that should be addressed. In particular, the "make or take" model for pricing exchange services has led to perverse outcomes. In the make or take model, trading platforms charge access fees to traders who "take" liquidity with marketable orders and pay rebates to limit order traders that "make" liquidity by placing standing limit orders. Current best execution standards require brokers to take the "best" price without regard to the access

Table 1 Fee Schedules US Exchanges

Venue	Take Fee ($/100 shrs)
AMEX	0.30
NYSE-Arca (ARCA)	0.30
CHX	0.30
DirectEdge X (EDGX)	0.30
Nasdaq (NDAQ)	0.30
BZX	0.29
NSX	0.29
NYSE	0.23
PSX	0.19
BYX	−0.02
EDGA	−0.04
Nasdaq OMX BX (BX)	−0.14
CBSX	−0.17

(Take) Fee schedules at US exchanges November 2012. Source: Battalio, Corwin, and Jennings (2016)

Table 2 Fee Schedule NYSE (extract)

LIQ	DESCRIPTION	TAPE A	
1	Taking Liquidity	$0.00275	$0.0030 if Adding ADV<250,000
1	Floor Take	$0.0024	
1	SLP Take	$0.00275	
1	DMM Take	$0.00275	
2	Providing	$(0.00140)	$(0.0000) If Non-Displayed
2	Providing Tier 3	$(0.00180)	Adding ADV>0.35% Tape A CADV and MocLoc>0.05%. $(0.0000) if Non-Displayed
2	Providing Tier 2	$(0.0020)	Adding ADV>0.75% Tape A CADV and MocLoc>0.10% $(0.0000) if Non-Displayed
2	Providing Tier 1	$(0.0022)	Adding ADV>1.10% Tape A CADV and MocLoc>0.12% $(0.0000) if Non-Displayed
2	Floor Provide	$(0.00190)	$(0.0022) if Floor provide >0.33% Tape A CADV, $(0.0020) if Floor provide >0.07% Tape A CADV, $(0.0022) if member is at providing Tier 1, $(0.0020) if Tier 2
2	SLP Provide	$(.0014/23)	Adding ADV> 0.20% Tape A CADV, otherwise (.0000/.0006) If Non-Displayed
2	SLP Tier 2	$(.0026)	Adding ADV > 0.45% Tape A CADV $(.0009) if Non-Displayed
2	SLP Tier 1A	$.(00275)	Adding ADV> 0.60% Tape A CADV, $(.00105) if Non-Displayed
2	SLP Tier 1	$(.0029)	Adding ADV> 0.90% Tape A CADV, $(.0012) if Non-Displayed

Fee structure at the NYSE, October 2016, Excerpt. Source: NYSE web cite.

> fees. We recommend that the SEC require that all brokers pass on the fees and liquidity rebates to their clients. The SEC also should indicate clearly that the principles of best execution apply to net prices and not to quoted prices. Alternatively, the SEC simply could ban access fees."

The problem is that retail investors tend to delegate trade routing decisions to brokers. The brokers typically charge fixed commisions to the retail investors. These brokers have incentives to send the orders to the venue offering the highest rebate, even if that does not offer the best price/limit order execution.

Is this a problem? Well, that is the subject of a study by Battalio et al. (2016). They show evidence that (some) brokers route their trades to the venues with the highest liquidity rebate. They also show evidence that this is not the best thing for the brokers to do on behalf of their customers. The venues with the highest rebates are not the ones giving the best execution. Essentially, they argue that their evidence show that Angel et al. (2011) are correct, that brokers can't have it all. If the broker chooses the highest fee venue, expect worse execution.

There is however alternative evidence for this. One example is Malinova and Park (2015), which looks at a change in maker/taker fees in Canada. They do not find the same problems.

The SEC is actively investigating whether maker/taker fees should be regulated further. The SEC Commission on Equity Market Structure Advisory Committee has recently (July 8, 2016) suggested an Access Fee Pilot study. In the pilot, they suggest the following different buckets, to which they plan to randomize 100 stocks (and ETFs) into

- (as now) 0.003 cents (Control group)
- 0.002 cents
- 0.001 cents
- 0.0002 cents (ie. almost zero)

a two year program is proposed, and one should investigate consequences on various aspects of liquidity and trading costs.

3.2.2 Payment for Order Flow

Payment for Order Flow Payment for Order flow is a practice where an exchange pays brokers a fee for routing certain kinds of orders to their exchange. The type of orders one typically want to attract are what in microstructure parlance is called *uninformed order flow*, which most often are the trades of retail investors. Exchanges like these orders, as they are less likely to be adversely selected.

The SEC defines payment for order flow as follows

> "As a way to attract orders from brokers, some exchanges or market-makers will pay your broker's firm for routing your order to them – perhaps a penny or more per share. This is called 'payment for order flow.' Payment for order flow is one of the ways your broker's firm can make money from executing your trade. The firm can also make money by internalizing your order.
>
> Upon opening a new account and on an annual basis, firms must inform their customers in writing whether they receive payment for order flow and, if they do, a detailed description of the type of the payments. Firms must also disclose on trade confirmations whether they receive payment for order flow and that customers can make a written request to find out the source and type of the payment as to that particular transaction." [SEC Webcite]

Payment for order flow is allowed in the US market, but not everywhere. The UK FSA has for example issued regulation outlawing the practice in the London market.

The reason for not allowing payment for order flow should be clear once you see the qualifications the SEC put on the practice. If retail traders are not aware of the practice, their brokers may choose to route trades

to the venue that maximizes payment for order flow, not venues that offers the best price (violation of the obligation to seek best execution).
Battalio and Holden (2001) argues that the driver of payment for order flow is the ability to identify characteristics of the trader. Maddoff, for example, only wanted retail order flow, as this was unlikely to be informed. Also, need to adjust concept of trade cost for broker commissions.

4 Measuring the cost of trading

As an investor, the most important information is the *cost* of trading.
Generally, two costs of trading

- Direct (fees) [already seen example: Maker/Taker fees]
- Indirect (market impact)

Second source of cost the most important.

4.1 Implementation Shortfall

Implementation Shortfall conceptually introduced by Perold (1988). Perold claim that to fully measure the cost of trading, need to account for two effects:

- the degree to which your own trading moves prices
- the probability that you are not able to buy (or sell) all or part of the portfolio (nonexecution risk)

His idea is to compare two value of two "paper portfolios"

- The desired portfolio using the prices extant when the decision to trade is made
- The actual portfolio one ends up with.

The difference of these two is the *implementation shortfall*.
To more formally define implementation shortfall, let us follow the notation in Foucault et al. (2013).
We start at time 0. At that time a trader decide to buy q of a given share. The portfolio position is to be evaluated at a future time t.
The desired portfolio, valued using prices available at trade initiation:

$$m_0 q$$

at time t this is worth

$$m_t q$$

This portfolio has a return of

$$R_p = q(m_t - m_0)$$

In actuality, to build up the portfolio will be trading at different prices in period from time 0 to t. Suppose the trader achieves a price $\overline{p}$ on the trades, but is only able to fill fraction $\kappa < 1$ of desired quantity q.
Then, the actual return on the part of the portfolio one gets is

$$R_a = \kappa q(m_t - \bar{p})$$

The implementation shortfall is the difference of these two

$$IS = q(m_t - m_0) - \kappa(m_t - \bar{p}$$

Rewriting this expression

$$IS \quad = \underbrace{\kappa q(\bar{p} - m_0)}_{\text{execution cost}} \quad + \underbrace{(1-\kappa)q(m_t - m_0)}_{\text{opportunity cost}}$$

This shows the decomposition of implementation shortfall into to components

- execution cost (prices change before on manages to buy)
- opportunity cost (one can not trade what one want)

The execution cost can be further decomposed into a part that is due to a delay between the order decision and the *implementation* decision (sending the order).
Suppose this happens at time $\tau \in (0, t)$. Then

$$\underbrace{\kappa q(\bar{p} - m_0)}_{\text{execution cost}} \qquad = \kappa q(\bar{p} - m_\tau) \quad + \underbrace{\kappa(m_\tau - m_0)}_{\text{delay portion}}$$

As these last calculations show, the information necessary to calculate implementation shortfall correctly is substantial. In practice we do not have this information. Estimates of implementation shortfall are therefore using various simplifications, such as assuming that missing quantities are traded at the closing price.

4.2 ITG - transaction costs estimates

ITG is a company providing services to "buy-side" investors.
It is best known for its report on trading costs.
We will look at these estimates.
Using data from trading by institutional investors, they estimate trading cost implied in these trades
The idea is to estimate the concept of *implementation shortfall*.
The usual way of calculating implementation shortfall.

$$\text{Implementation Shortfall} = \begin{cases} \text{Trade Price} - \text{Benchmark Price} & \text{for a buy order} \\ \text{Benchmark Price} - \text{Trade Price} & \text{for a sell order} \end{cases}$$

What ITG reports is broken down estimates of these.

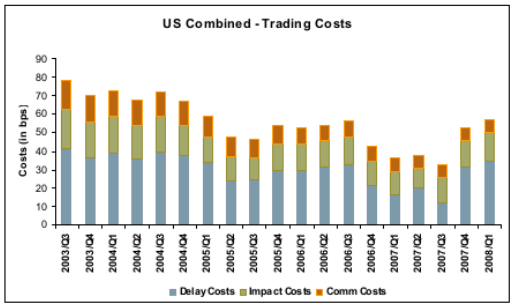
$$\text{Total Cost} = \text{Implementation Shortfall} + \text{Commisions}$$

4.2.1 Trading costs for the US, aggregate

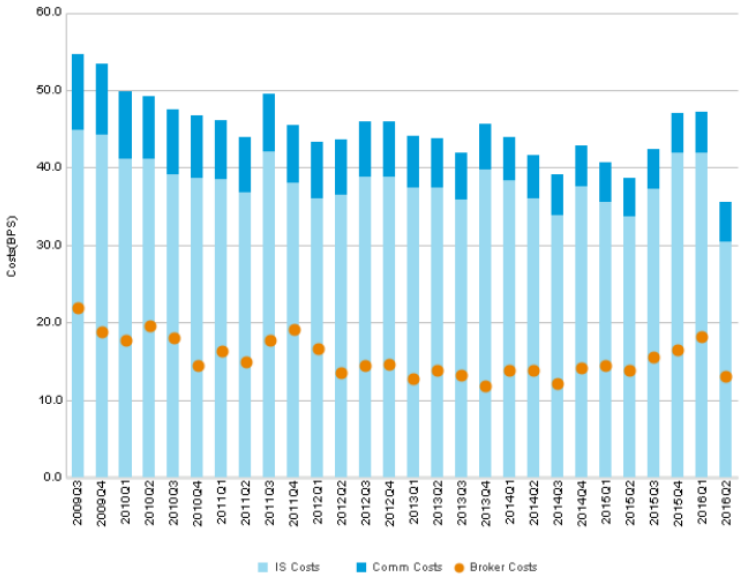
Let us look at the case of averages for the US. ITG has been doing estimation since 2004. Figure 20 shows their estimates from 2004 till 2016.
Let us also look at the numbers, shown in table 3 One thing of note is obviously magnitudes of trading cost. The total costs vary between 80 and 36 bps. I.e. when trading, one expects to pay between 0.4% and 0.8% to make a trade. Another interesting observation is the time-trend. From 80 basis points in 2004 to 36 in 2016.

Figure 20 Trading costs estimates 2008 and 2016 report US Combined

Panel A: 2008 report



Panel B: 2016 report



Source: ITG (2016).

Table 3 Trading costs estimates 2008 report US Combined

Panel A: 2008 report

Estimates of trading costs, US Combined (all costs in bps)

	Delay Costs	Impact Costs	Comm Costs	Total Costs
2003/Q3	42	21	16	79
2003/Q4	36	19	14	70
2004/Q1	39	20	14	73
2004/Q2	36	19	14	68
2004/Q3	39	20	13	72
2004/Q4	37	17	12	67
2005/Q1	34	14	11	59
2005/Q2	24	13	10	47
2005/Q3	24	12	10	47
2005/Q4	29	15	10	54
2006/Q1	30	14	9	53
2006/Q2	32	14	8	54
2006/Q3	33	15	9	57
2006/Q4	22	13	8	43
2007/Q1	16	13	7	36
2007/Q2	20	11	7	38
2007/Q3	12	14	7	33
2007/Q4	32	15	7	53
2008/Q1	34	16	7	57

Panel B: 2016 report

United States(BPS)

Quarter	IS Costs	Comm Costs	Total Costs	Broker Costs
2009Q3	44.9	9.7	54.6	22.0
2009Q4	44.3	9.1	53.4	18.8
2010Q1	41.1	8.8	49.9	17.7
2010Q2	41.1	8.2	49.3	19.6
2010Q3	39.2	8.4	47.3	18.1
2010Q4	38.7	8.1	46.7	14.4
2011Q1	38.6	7.5	46.1	16.4
2011Q2	36.8	7.2	43.9	14.9
2011Q3	42.1	7.5	49.6	17.7
2011Q4	38.0	7.5	45.5	19.1
2012Q1	36.0	7.3	43.2	16.6
2012Q2	36.5	7.2	43.7	13.6
2012Q3	38.9	7.1	46.1	14.5
2012Q4	38.8	7.2	45.9	14.6
2013Q1	37.5	6.6	44.1	12.7
2013Q2	37.5	6.3	43.7	13.8
2013Q3	35.9	6.1	42.0	13.3
2013Q4	39.8	5.8	45.5	11.9
2014Q1	38.3	5.6	43.9	13.8
2014Q2	36.1	5.6	41.7	13.8
2014Q3	33.8	5.4	39.2	12.1
2014Q4	37.6	5.3	42.9	14.2
2015Q1	35.6	5.0	40.6	14.5
2015Q2	33.7	5.1	38.7	13.8
2015Q3	37.3	5.1	42.4	15.5
2015Q4	42.0	5.1	47.1	16.5
2016Q1	41.9	5.4	47.3	18.3
2016Q2	30.4	5.1	35.6	13.1

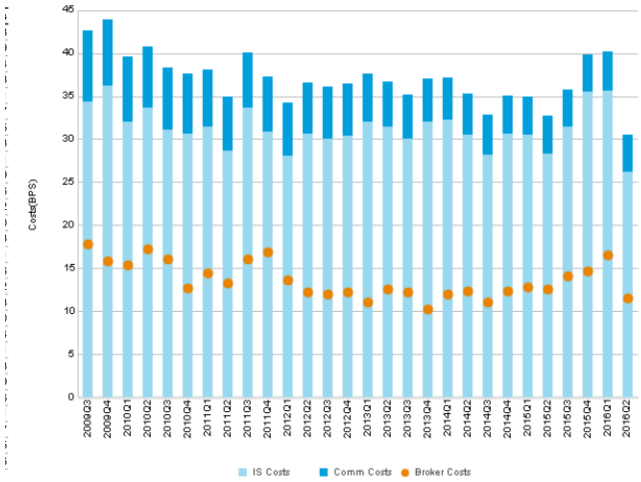
Source: ITG (2016).

4.3 Linking trading costs to type of stock

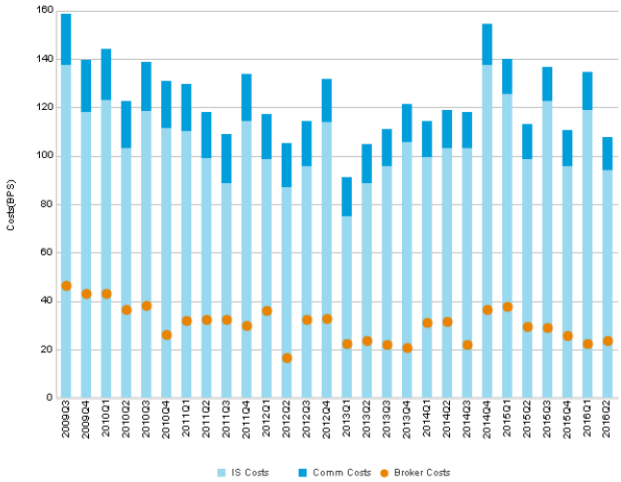
Let us now look at trading costs broken down on different types of stocks. Figure 21 show the 2016 report estimates. Smaller cap stocks have much higher estimated costs.

Figure 21 Trading costs estimates 2016 report US Large Cap and Small Cap

Panel A: Large Cap



Panel B: Small Cap



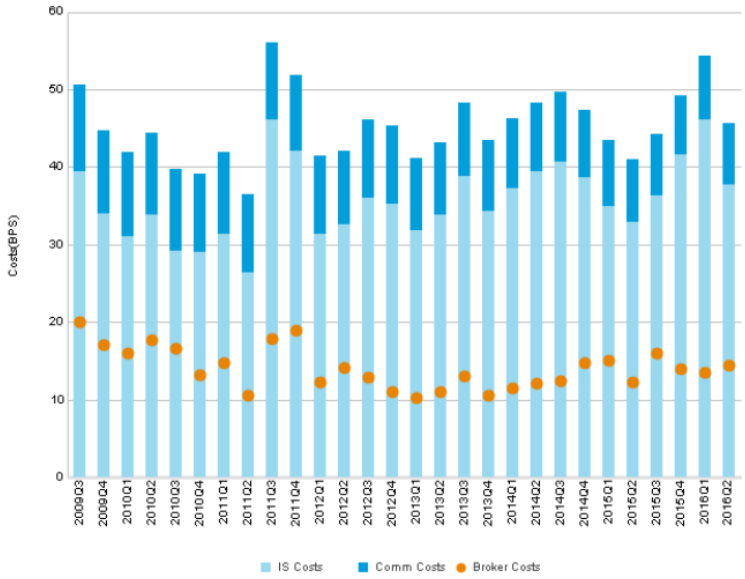
Source: ITG (2016).

4.4 European costs

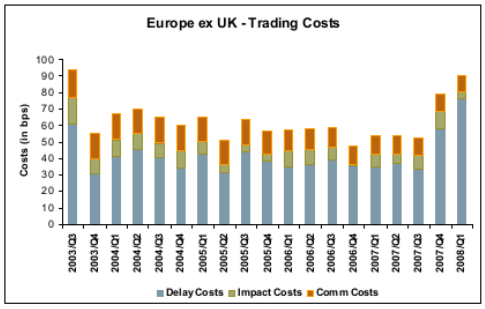
Let us also look at the case of Europe (ex UK)

Figure 22 Trading costs estimates Europe (ex UK)

Panel A: 2016 Report



Panel B: 2008 Report



Source: ITG (2016).

4.4.1 Longer term trends, US

Let us look at the evolution over time with an even longer perspective, and turn to some academic estimates. The ITG estimates are attempts to measure trading costs, taken from actual trades.

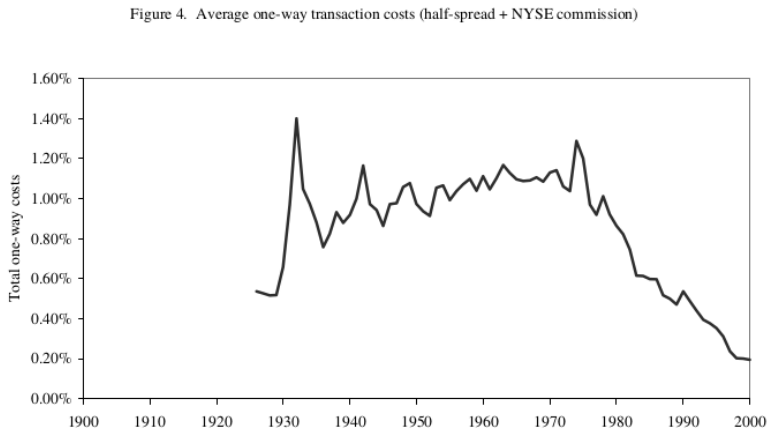
Academics do not have access to that kind of data, and instead measure similar cost measures taken from the record of trades at exchanges.

- Bid/Ask Spread (in percent of price). The theoretical cost of a round trip. Going back in history observed at the end of the day.

- Effective spread. Close to the price impact calculation. The difference between a transaction price and the midprice just before the trade took place (the "unperturbed" price).

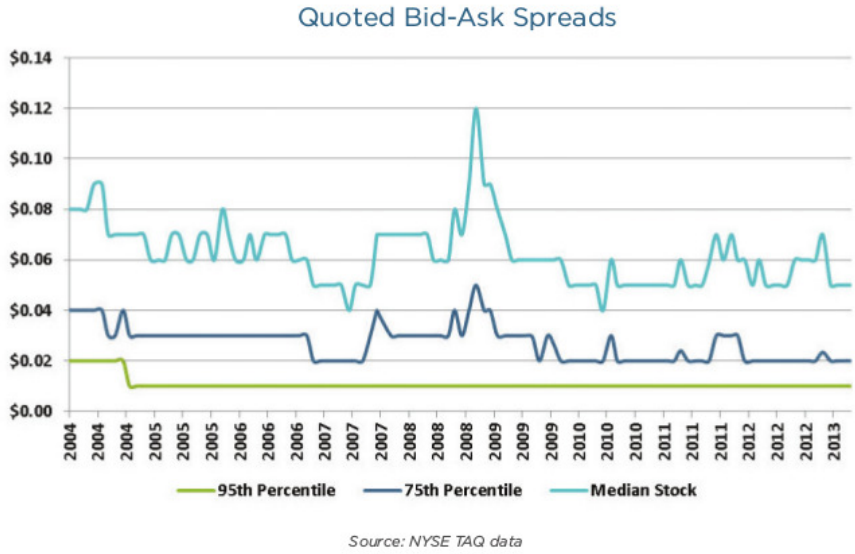
Show some estimates of both. Let us first look at the US in the last century. Figure 23 shows the effect of the deregulation of trading in the nineteen eighties in the US.

Figure 23 Average Transaction Costs Dow Jones Stocks, 1900–2000



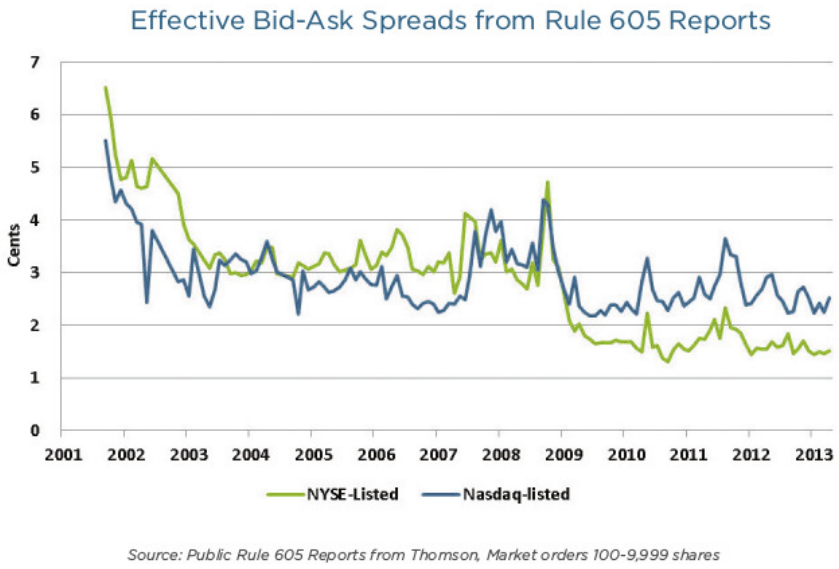
Source: Jones (2002).

Figure 24 Quoted Bid-Ask Spreads



Source: Angel et al. (2015).

Figure 25 Effective Bid-Ask Spreads

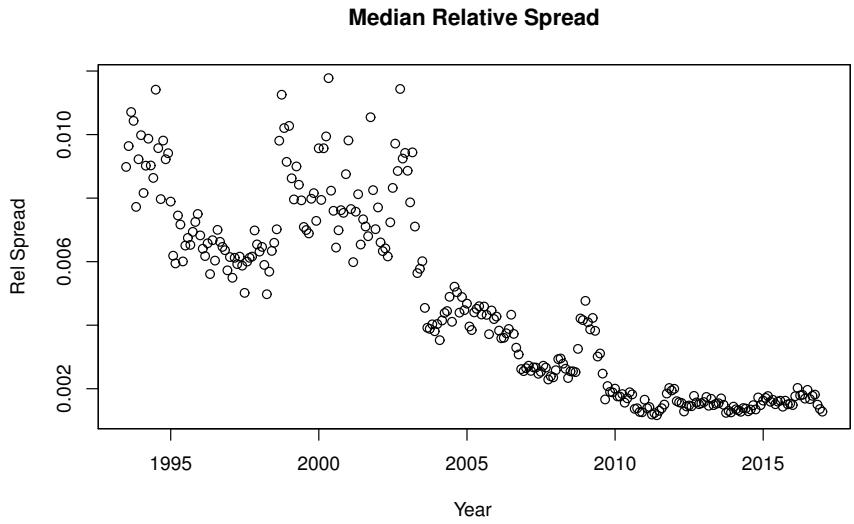


Source: Angel et al. (2015).

4.4.2 Longer term trends, Norway

Figure 26 shows closing spreads for the OBX constituents in the period after 1993.

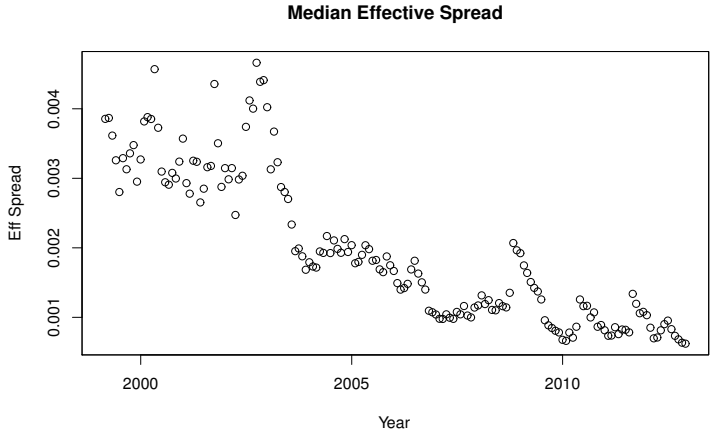
Figure 26 Time Series, Closing Relative Bid/Ask Spread, Monthly Median, OBX constituents



Time series plot of median relative spread across stocks on the Oslo Stock Exchange.The sample is restricted to stocks in the OBX index, an index of the 25 most active stocks on the OSE. The relative spread is the difference between closing bid and ask prices, divided by the last trade price. Actual numbers (not percentages).

Figure 27 shows estimates for effective spreads for the same sample of stocks.

Figure 27 Time Series, Effective Spreads, Monthly Median, OBX constituents



Time series plot of effective spreads across stocks on the Oslo Stock Exchange. The sample is restricted to stocks in the OBX index, an index of the 25 most active stocks on the OSE. The effective spread is the difference between a trade price and the midquote just before the trade, relative to the midquote. Actual numbers (not percentages).

4.4.3 Concluding so far

Essentially showing that estimates of trading costs for a random trade have fallen everywhere.

4.4.4 But what about larger traders/institutions?

Problem: Institutional traders trade larger quantities. Worry more about price impact.
Also worry about frontrunning by other traders: If you want to fill a large quantity, and other traders become aware of this, incentives for others to "get in ahead of you" (pinging).
So the cumulated market impact may be larger for institutions. (Main worry).
Active (and unsolved) research issue. To what extent is this something that institutions should worry about?
Academics are actively looking for a smoking gun here, still haven't found it.
Still, this effect does need to be rather large to offset the generally lower transaction costs in the current market place(s).

What about the very large trades? Larger trades have always been hard.
Old Days: Most Exchanges had special facilities for dealing with "blocks."
(Upstairs Market)
NYSE: Block Desk a separate mediated market, brokers using phones to trusted counterparties (which should not use the information about the block in the downstairs market).
Final price of a block trade usually contain a premium/discount relative to market price.
Block trading replaced with algorithms trickling the block into the continous limit order book.
If we want to make comparisons need to ask what is the aggregated price impact of this "trickled" trade, and how does this compare to the typical premium/discount for the same sized block in the "old days"

Kraus and Stoll (1972) estimates The paper by Kraus and Stoll (1972) is a famous case study of block trades in New York. The question it tries to answer is: What is the price impact you need to allow for when buying (or selling) a *large* position in a given stock. Numbers such as spreads, be they quoted, effective, or what have you, can be argued to be estimates of the trading costs of a *small* position, as it measures the cost if you trade within the current quotes in the limit order book. If you try to trade a *large* position, you will exhaust the buy (or sell) interest in the book. While others may appear to replenish the order book, these need to be compensated by a higher (if buying) or lower (if selling) price.
The estimates in Kraus and Stoll are based on comparing the price in a block trade (which was agreed off the floor) with the current price. The definition of a block is fluid, but typically, trades above 10 thousand shares is viewed as a block. This difference is an estimate of the price concession one need to make to trade the whole block. Table 4 summarizes the necessary price concessions. If we measure relative to previous day's close, selling (buying) a block would mean a price concession of -1.86% ($+1.50\%$). If we alternatively measure relative to the price on the floor when the block price is agreed, the concession is smaller in absolute value. However, some of this price movement can be due to information leakage, it information about the block being negotiated upstairs also reaches the floor... So it is not obvious what measure to use.

Table 4 Price Concessions

	Relative to previous days close	Immediate price impact	Temporary (reversal)
Sells	1.86%	1.14%	0.71%
Buys	1.50%	0.75%	0.09%

From Kraus and Stoll (1972).

The best ways to figure out what is going on is to look at the pictures from the paper, which summarizes

the most salient results, which is reproduced in figures 28 and 29 below

Figure 28 Estimates of price impact of a block sale

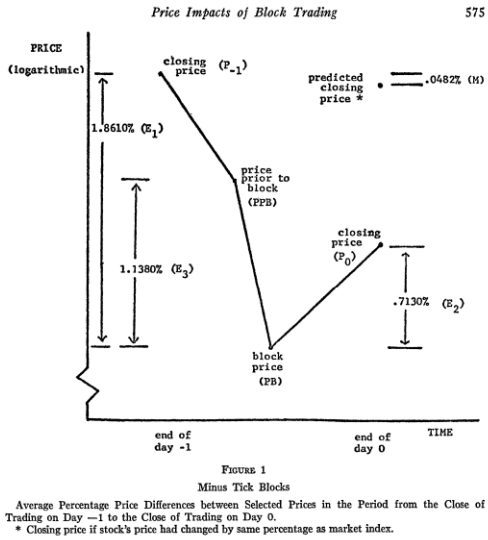


Figure taken from Kraus and Stoll (1972).

Figure 29 Estimates of price impact of a block buy

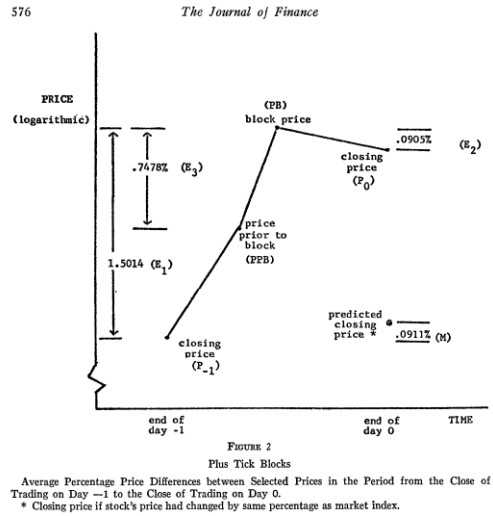
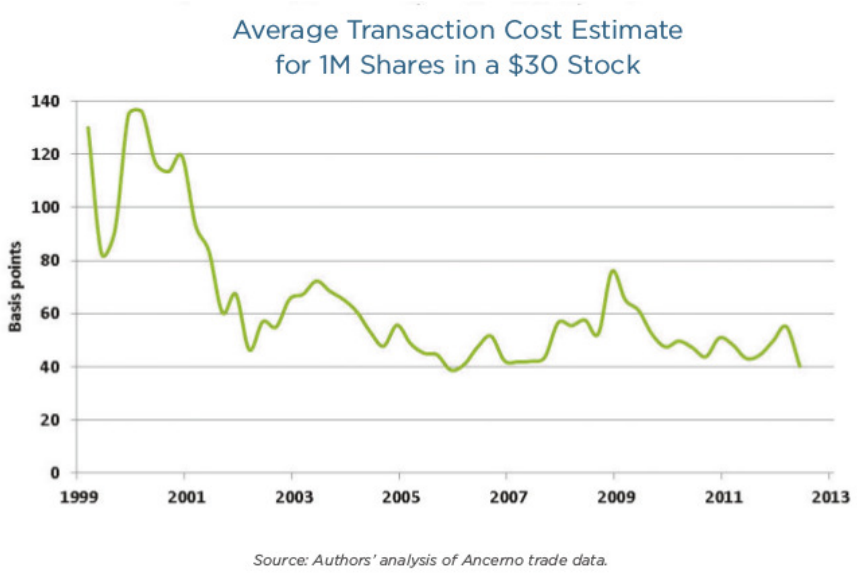


Figure taken from Kraus and Stoll (1972).

Let us for comparison show estimates for block trades in recent years, taken from Angel et al. (2011), in figure 30.

Figure 30 Transaction Costs - Block Trades



Source: Angel et al. (2015).

5 Regulators to the rescue?

Financial Regulation a growth industry

Prone to over-regulating (Dodd-Frank-2300-pages)

But some regulation important for traders: The emphasis on trade reporting, also on statistical reporting by trade venues of actual transaction costs.

So for example, for the US, any exchange need to report estimates of the cost of trading at that venue.

Securities And Exchange Commission: Rule – Disclosure of Order Execution and Routing Practices – Better known as Rule 605 Reports. – Problem: Exchanges gaming this by not making it particularly easy to access the data .... and the data they make it easy to find tend to be complementary to that particular market place

See for example the two pictures take from the NYSE's summary of their market quality report, which looks more like pure promotional material for the NYSE. (Figures 31 and 32.)

Figure 31 Market Share NYSE

Source: NYSE Market Quality Report, 2016. (NYSE's summary of the material that they report to the SEC.)

But still, investors should be asking their brokers for the basis for choice of trading venue. How are the execution quality numbers at that venue?

Also, regulation cares about the concept of "best execution", both MiFiD and Reg-NMS. But usually ambigous.

Figure 32 Time at Best NYSE

Source: NYSE Market Quality Report, 2016. (NYSE's summary of the material that they report to the SEC.

6 Lessons for investors

The important thing is always to buy the right stocks and sell the wrong stocks. However, the lesson from this talk is that it matters how you *implement* those buying and selling decisions.

Trading costs — a Drag

Not watching trading costs can be compared to ignoring biofouling for shipping.

Biofouling

Biofouling or biological fouling is the accumulation of microorganisms, plants, algae, or animals on wetted surfaces. (Wikipedia)

Illustration (magnified) of ships hull that's been in the water a while:

Bifouling a major problem for shipping

- Reduce speed.

  In the day of the sailing ship bifouling reduced speeds up to a third.

  Even with todays motorized ships

  - speeds can be reduced up to 10%.
  - $\rightarrow$ increased fuel consumption

- Unprotected surfaces are gradually destroyed

Minimize the Drag! Things to remember

- Transaction costs may put such a drag on a strategy that it should be reconsidered.
- High Frequency Trading not necessarily your enemy
- What are the available trading venues?
- What is the fee structures in a given venue (e.g. make/take fees)
- What is the liquidity (availability of quantity) in a given venue.
- There is a tradeoff pre-trade price information (public venue) and hiding quantities (dark pools).
- The importance of asking pointed questions to brokers and intermediaries about trade choices.

  Why that particular venue/order type/method?

7 Gazing into the Crystal Ball

What can we expect to happen?

- We are not going to go away from electronic trading.
- Expect continued process of
  - mergers of trade venues
  - new trade venues
  - innovation in ways of trading
- But:
  - There *are* economics of scale in the trading industry.
  - At some point we may see concentration of trading in larger venues.
- Regulation is a driver of change in this industry
  - Much regulation it is the outcome of political processes.
  - Have you looked at US/UK/European politics lately? (Brexit, the Donald, ...)

8 Further Reading

A good overview of microstructure from the point of view of investors: Harris (2015)
Some perspectives on the recent fifteen years in the US markets: Angel et al. (2011, 2015)

References

James J Angel, Lawrence Harris, and Chester S Spatt. Equity trading in the 21st century. *Quarterly Journal of Finance*, 1(1): 1–53, 2011.

James J Angel, Lawrence Harris, and Chester S Spatt. Equity trading in the 21st century: An update. *Quarterly Journal of Finance*, 5(1), 2015.

Robert Battalio and Craig W Holden. A simple model of payment for order flow, internalization, and total trading cost. *Journal of Financial Markets*, 4:33–71, 2001.

Robert Battalio, Shane A Corwin, and Robert Jennings. Can brokers have it all? On the relation between make-take fees and limit order execution quality. *Journal of Finance*, LXXI:2193–2238, 2016.

Bruno Biais. High frequency trading. Presentation prepared for the European Institute of Financial Regulation, September 2011.

Thierry Foucault, Marco Pagano, and Ailsa Roell. *Market Liquidity. Theory, Evidence and Policy.* Oxford University Press, 2013.

Larry Harris. Trading and electronic markets: What investment professionals need to know. CFA Institute Research Foundation, September 2015.

ITG. Global cost review q2/2016. Working Paper, ITG (Investment Technology Group, October 2016.

Charles M Jones. A century of stock market liquidity and trading costs. Working Paper, Columbia University, May 2002.

Kjell Jørgensen, Johannes Skjeltorp, and Bernt Arne Ødegaard. Throttling hyperactive robots - Order to trade ratios at the Oslo Stock Exchange. Working Paper, University of Stavanger, January 2017.

Alan Kraus and Hans R. Stoll. Price impacts of block trading on the New York Stock Exchange. *Journal of Finance*, 27(3): 569–588, 1972.

Gregory Laughlin, Anthony Aguirre, and Joseph Grundfest. Information transmission between financial markets in chicago and new york. *Financial Review*, 49:283–312, 2014.

Michael Lewis. *Flash Boys*. Allen Lane, 2014.

Katya Malinova and Andreas Park. Subsidizing liquidity: The impact of make/take fees on market quality. *Journal of Finance*, LXX(2):509–537, April 2015.

Andre Perold. The implementation shortfall: Paper versus reality. *Journal of Portfolio Management*, 14(3):4–9, 1988.