Peter Gomber - Satchit Sagade - Erik Theissen - Moritz Christian Weber - Christian Westheide

Competition/Fragmentation in Equities Markets: A Literature Survey

SAFE Working Paper Series No. 35
Non-Technical Summary

Financial markets in the US and Europe have evolved from a largely consolidated structure comprising the primary listing venues, regional exchanges and the OTC markets, to a highly competitive but fragmented structure comprising several exchanges and alternative trading systems like crossing networks, electronic communications networks, dark pools, and multiple broker-dealers offering internalization services. These changes have been fueled by an increase in computational capabilities, improvements in networking technologies and several regulatory initiatives like the Regulation-National Market System (Reg-NMS) in the US and the Markets in Financial Instruments Directive (MiFID) in the EU. While these changes have brought about several benefits like reduced transaction costs, regulators and market participants have raised concerns about the potential adverse effects associated with an increase in execution complexity as well as the impact on market quality of new types of venues like dark pools. In this article we have reviewed the theoretical and empirical literature, which examines the reasons why markets fragment as well as the resulting impact on different aspects of market quality.

The literature surveyed provides several important lessons: First and foremost, the reasons underlying market fragmentation need to be examined keeping in mind two opposing economic forces acting simultaneously in securities markets. On the one hand, due to the presence of strong network externalities, financial markets have a tendency to consolidate in space and time. Additionally, there are also economies of scale in concentrating order flow, as it allows exchanges to amortize fixed costs over more trades. On the other hand, given the heterogeneity of traders' preferences it is difficult to envisage a single market that can cater to the needs of all market participants - traders differ with respect to their motives, required levels of transparency or opacity, order sizes, and levels of patience.

Second, while examining the larger welfare implications of market fragmentation, it is important to recognize that embedded in securities markets is the market for trading services as well as the market for the traded securities. While the reduced explicit transaction costs resulting from competition in the market for trading services can be beneficial to market participants, competition in the market for the traded securities can be harmful as the quality of prices may be adversely affected potentially increasing implicit transaction costs. The theoretical and empirical literature examining the impact of dark, non-transparent venues seems to be a case in point.

Recent developments in the market structure of equities and other asset classes have raised several questions which remain unanswered. The reasons as to why different stocks fragment differently are still unclear. For example, according to LiquidMetrix (2010) London Stock Exchange's (LSE) market share in FTSE 100 stocks ranged from 50% to 80% whereas the market share of the BATS Exchange ranged from 2% to 15%. Similar patterns can also be observed in continental European equities and stocks traded across the Atlantic. Empirical studies have also shown that the market share of the OTC market is significant – in Europe it is around 40% for
equities as per Gomber and Pierron (2010). The reasons for these patterns in fragmentation also remain unclear. Last, but not the least, as compared to equities, research examining the impact of fragmentation/competition in other asset classes is scant, and in some cases nonexistent. This question is especially important in light of the MiFID II proposals currently being debated in Europe, under which a significant portion of OTC trading in all asset classes will be subjected to increased transparency requirements and regulatory oversight.

The complex, subtle and simultaneous interactions of the forces discussed in this article make the task of designing optimal policy responses and regulatory interventions extremely difficult. Policy-makers, regulators and exchanges should carefully evaluate the impact of their decisions keeping in mind these issues.
Competition/Fragmentation in Equities Markets: 
A Literature Survey

Peter Gomber∗  Satchit Sagade†  Erik Theissen‡  
Moritz Christian Weber§  Christian Westheide¶

November 2013

Abstract
Advances in technology and several regulatory initiatives have led to the emergence of a competitive but fragmented equity trading landscape in the US and Europe. While these changes have brought about several benefits like reduced transaction costs, regulators and market participants have also raised concerns about the potential adverse effects associated with increased execution complexity and the impact on market quality of new types of venues like dark pools. In this article we review the theoretical and empirical literature examining the economic arguments and motivations underlying market fragmentation, as well as the resulting implications for investors’ welfare. We start with the literature that views exchanges as natural monopolies due to presence of network externalities, and then examine studies which challenge this view by focusing on trader heterogeneity and other aspects of the microstructure of equity markets.

Keywords: Market Structure, Competition, Fragmentation, Liquidity, Market Quality
JEL Classification: G10

∗Chair of e-Finance, Goethe University, +49 69 798 34682, gomber@wiwi.uni-frankfurt.de
†Center for Financial Studies and Center of Excellence SAFE, Goethe University, +49 69 798 30085, sagade@safe.uni-frankfurt.de
‡Finance Area, University of Mannheim and Center for Financial Studies, Goethe University, +49 62 118 11517, theissen@uni-mannheim.de
§Chair of e-Finance, Goethe University, +49 69 798 33854, moweber@wiwi.uni-frankfurt.de
¶Finance Area, University of Mannheim and Center for Financial Studies, Goethe University, +49 62 118 11822, westheide@uni-mannheim.de

We gratefully acknowledge research support from the Center of Excellence SAFE, funded by the State of Hessen initiative for research LOEWE.
1. Introduction

Since the implementation of Regulation-National Market System (Reg-NMS) in the US and Markets in Financial Instruments Directive (MiFID) in the EU, equities market structure on both sides of the Atlantic has undergone a remarkable change. Undoubtedly, one of the successes of these regulatory initiatives has been the emergence of a truly competitive marketplace. O’Hara and Ye (2011) note that in 2008 there were 10 exchanges, 5 electronic communication networks, 20 or more alternative trading systems and more than 100 broker-dealer firms offering internalization services in the US. In the EU there were 22 regulated markets, 27 multilateral trading facilities, 23 dark pools and 9 broker-dealer crossing networks as of June 2010 (Gomber and Pierron (2010)). While these changes have led to significant improvements in market quality, regulators and market participants have also raised concerns on whether unintended consequences like proliferation of dark trading venues, extreme fragmentation of order flow and emergence of high-frequency trading have really helped financial markets achieve their core objectives of efficient capital allocation, risk transfer and price discovery. Commenting on the US experience, Tabb (2013) observes that “…Reg. NMS started a progression of technology changes that has exacerbated execution complexity, crippled the NYSE floor, prioritized speed over liquidity, and virtually put a bullet in the head of smaller, less-technology-adroit brokers and floor traders. It also fragmented the markets, drove ever-increasing messaging rates, created order-type complexity, and arguably enabled high-frequency traders to take advantage of the very investors Reg. NMS was intended to protect, while actually making the markets less transparent for regulators.” In light of these concerns, fresh regulatory initiatives have already been proposed or are currently being debated. As these initiatives will most certainly define the direction taken by world equity markets over the next decade, it is imperative that they are grounded in economic theory and evidence.

Microstructure economists have always been interested in the consolidation versus competition debate. One of the earliest references can be found in Stigler (1964) who argues that “The performance of the main function of the exchange as a market-place is subject to economies of scale. The greater the number of transactions in a security concentrated in one exchange, the smaller the discontinuities in trading and the smaller

1Angel et al. (2010) provide a detailed overview of the achievements of the regulatory initiatives implemented in the US over the last decade.

the necessary inventories of securities. As a result the price of a security will almost invariably be “made” in one exchange.” In the 1970s, as part of the proposals to introduce a National Market System (NMS), the Securities and Exchange Commission (SEC) proposed to repeal Rule 394 of the NYSE to encourage competition between exchanges and the so-called third and fourth markets. This proposal was the subject of considerable debate among academics, policy-makers and market participants. Smidt (1971), arguing in favor of competition in the provision of market-making services, called for the rule to be repealed. He argued that exposing NYSE specialists to effective competition would increase the supply of liquidity. On the other hand, Bloch and Schwartz (1978) argued that in the securities markets the market for the provision of trading services accompanies the market for the traded securities; while increased competition would benefit investors by reducing the explicit costs of trading, the resulting fragmentation would damage the quality of prices. They further argued that an increase in order flow internalization by brokers would unleash a vicious circle where order flow on the NYSE would dry up leading to its eventual demise. Hamilton (1979) examined this question empirically by regressing NYSE spreads and return volatility on the proportion of total volume executed off-board and the level of competition in each stock, controlling for several variables like stock price, institutional interest and level of informed trading. He concluded that the benefits of competition outweighed the costs, and hence argued in favor of regulatory intervention that would increase competition. After an unsuccessful attempt in 1977, the SEC finally repealed Rule 394 by adopting Rule 19c-3 which allowed off-board trading with effect from July 18, 1980.

In this article we review the theoretical literature and empirical evidence on both sides of the debate. In the process, we also chart the evolution of equity markets, starting from floor-based trading concentrated on major exchanges in the 1960s and 1970s, to a highly competitive, technologically-advanced marketplace with multiple lit and dark venues allowing investors to trade at speeds approaching the speed of light. While a brief review of the literature on this topic can be found in O’Hara (1995), Madhavan (2000) and de Jong and Rindi (2009), our approach differs from these surveys in two aspects. First, by exclusively focusing on the consolidation versus fragmentation debate, we are able to provide a more detailed analysis of the literature as opposed to the

---
3NYSE Rule 394 (and AMEX Rule 5) prohibited member firms from trading listed securities off the exchange floor without the exchange’s permission even if they could find a better price. Several institutions responded to this restriction by trading in the ‘third market’ with non-member firms, and, after the introduction of Instinet in 1969, the ‘fourth market’ comprising firms connected to each other through Instinet’s communication networks.

4Appendix A lists some of the important milestones in the evolution of the equity market structure.
above reviews which discuss this debate within the larger context of the microstructure of financial markets. Second, all the existing surveys are dated and as such do not cover the recent literature which focuses on some of the changes in the equity market structure discussed above. In the interest of brevity, we do not attempt to incorporate all the papers relevant to the topic. Instead, we examine each aspect of this debate by focusing on the contribution of the most important papers, while also highlighting important policy issues relevant to this debate.

We start by looking at the first attempts to study competition in financial markets in the next section. In Sections 3 we examine the first wave of studies which viewed exchanges as natural monopolies due to the presence of network externalities and economies of scale. In Section 4, we look at how the nature of competition in equities markets changed over the last 30 years. In Section 5 we review the literature examining the role of trader motives in market fragmentation. Here we look at the cream-skimming hypothesis, which investigates competition for informed and uninformed order flow and its adverse consequences, and trading in the upstairs and over-the-counter markets, which allow traders to efficiently execute large blocks. In Section 6 we examine competition between venues when the primary listing exchange operates as a limit order market with or without the presence of specialists or market-makers. In Section 7 we review the literature on market transparency focusing on the impact of different levels of pre-trade and post-trade transparency on competition among exchanges and alternative trading systems. We also examine the literature analyzing the controversial role played by dark pools in a competitive marketplace, and recent empirical studies which focus on the potentially opposite impact of dark and lit fragmentation on market quality and investor welfare. The impact of multiple tick size regimes and maker-taker pricing is the focus of Section 8 and finally we conclude in Section 9.

2. The First Attempts

Garbade and Silber (1978, 1979) examine the short-term behavior of prices in a market where trading is fragmented due to positive transportation and/or communication costs. The authors argue that the level of price integration between venues and the speed with which prices converge is inversely related to these costs. Furthermore, prices can adjust symmetrically such that the prices in both markets move towards each other. Alternatively, there can be a one-sided adjustment where the price in one market moves towards the price in the other market (dominant-satellite market). Based on an empirical examination of prices in the New York Stock Exchange on one hand, and in the Midwest
and Pacific Exchange on the other, the authors find that the relationship between them is that of a dominant-satellite market with the NYSE playing the role of the dominant market. The authors also observe that the introduction of new technologies that reduced communication delays over the previous century led to more integrated markets.

Cohen et al. (1982) analyze the operating efficiency of, and investor welfare in, a market comprising a central limit order book (without the presence of specialists), and a broker-dealer, who operates an independent limit order book and offers internalization services to his customers (with or without assuming a dealerships role). The authors argue that internalization is profitable to brokers as it allows them to queue-jump the central limit order book and avoid sharing commission with the exchange. They further observe that fragmentation leads to poorer market performance because executions on the exchange violate price priority leading to reduction in order flow routed to the exchange. The authors conclude that, if appropriate side payments like commission-sharing or exposing best in-house quotes to the market are not put in place, fragmentation can be harmful for market quality. However, Stoll (1982) questions the conclusions of Cohen et al. (1982) by arguing that exchanges are natural monopolies and investors will choose brokers who send orders to the most active venue. An internalizing broker-dealer who queue-jumps the primary listing venue in a fragmented market would only observe a fraction of the total order flow and hence will not be able to immediately match customer market orders with her limit orders. This will increase the expected time to execution, thereby eliminating the incentive for a customer to use such a broker in the first place.

Mendelson (1987), using the underlying framework provided by Amihud and Mendelson (1980, 1982), compares the operational characteristics of a market under four alternative market structures: a consolidated call auction market, a fragmented call auction market, a consolidated dealer market and a fragmented dealer market. The call auction market generates a single market clearing price whereas the dealer market is one where dealers post quotes for others to trade at. The impact of these alternative structures on total volume, price variance, quality of price signals and gains from trade are compared with the implementation costs associated with each of them. The authors conclude that in case of the call auction market, consolidation provides higher volume, less price volatility and higher gains from trade but less price efficiency as compared to fragmentation, while the results are generally ambiguous in a dealer market.

While these earlier studies do shed light on some of the issues central to the fragmentation versus consolidation debate, the approach used in these studies for the most part is largely informal (with the exception of Mendelson (1987)). Empirical studies also largely ignore important econometric considerations. For example, Hamilton (1979),
while acknowledging the potentially endogenous relationship between on-exchange bid-ask spreads and off-board trading, completely ignores it in his analysis.

3. Exchanges as Natural Monopolies

In this section, we review studies which view exchanges as natural monopolies due to the presence of strong network externalities and economies of scale. Traders obtain a better execution in a consolidated market as more traders are present in such a market leading to higher liquidity. This idea was formally developed in a series of papers by Pagano (1989a,b), Chowdhry and Nanda (1991) and others.

Pagano (1989a) examines the properties of the equilibria in fragmented markets by focusing on the feedback loop between trading volume and liquidity. The author considers a two-period game-theoretic model in which a set of traders with different endowments choose their demand to maximize a mean-variance utility function of terminal wealth. In the case where each trader chooses to trade on one of two centralized markets which are exactly similar, the resulting equilibrium is unstable such that a small change in trader expectations can lead to concentration of trading on one exchange and the disappearance of the other exchange. A stable separating equilibrium exists only if one market has a higher transaction cost accompanied by a higher absorptive capacity (more number of traders and greater cross-sectional diversity) such that large traders are then willing to trade on such a market. The author also considers a scenario where a trader can choose to trade on a single centralized exchange or search for a trading partner in the over-the-counter market for a fixed cost and uncertain outcome. In this case the author finds that the two markets can coexist and fragmentation is not necessarily welfare-reducing.

Chowdhry and Nanda (1991) extend the Kyle (1985) framework to a security that trades in multiple markets in order to examine the strategies of informed and uninformed traders. Their model consists of an informed trader, a market-maker, and two groups of liquidity traders: one large liquidity trader who takes the effect of her trading on prices into account and splits her trades across markets, and small liquidity traders each of whom trade small quantities in a single market. Just as in Kyle (1985), all traders are assumed to be risk-neutral, and each market-maker observes the consolidated order flow in her market and sets prices using a linear pricing rule in order to attract order flow.

Network externalities are the effects on a user of a product or service, of others using the same or compatible products or services. For example, social networks like Facebook are more valuable to each user if there are more users.

This result is similar to Admati and Pfleiderer (1988) who also analyzes the feedback loop between trading volume and liquidity to argue in favor of temporal concentration in trading.
When the small liquidity traders do not have discretion over their choice of market, the authors observe that prices become more informative as the number of markets increase. This is because small liquidity trades are uncorrelated across markets and due to this the diversification effect arising from aggregating all their trades in one market is lost which allows the informed trader to trade more aggressively. However, when a portion of small liquidity traders can choose their market, in equilibrium all of them choose the market with the lowest transaction costs. The same market also gets the largest share of trades by the large liquidity trader as well as the informed trader and in the limiting case where the number of non-discretionary small liquidity traders goes to zero, all trading is concentrated in a single market.

Empirical studies like Garbade and Silber (1979) and Hasbrouck (1995) also find that even when multiple markets coexist, the primary market is generally responsible for the bulk of the price discovery. For example, Hasbrouck (1995) observes that for NYSE stocks traded on the regional exchanges, NYSE is responsible for more than 90% of the price discovery. This tendency of fragmented markets to coexist for long periods of time despite strong consolidating forces was later termed as the network externality puzzle by Madhavan (2000). In such a fragmented marketplace there is always a worry that the benefits of consolidation discussed above are lost.

Harris (1993) argues that markets fragment because traders are heterogeneous with respect to their level of sophistication and motives to trade, and hence different market structures better serve the needs of some traders than others. For example, traders differ on the size of their trades, on whether they are information-driven or liquidity-driven, and on how patient they are while trading. The benefits of consolidation can also be achieved in fragmented markets if information is allowed to flow freely among different segments. The author defines such a market as a segmented market. In the remainder of the article, our focus will be on studies that try to generate models with stable equilibria by analyzing these differences. However, we first look at how the nature of competition in equity markets changed over the last few decades.

4. Competition in Equities Markets

In the 1980s and 1990s, regional exchanges like the Philadelphia Stock Exchange and the Pacific Stock Exchange traded stocks listed on the NYSE and NASDAQ pursuant to

---

7In a more recent study, O’Hara and Ye (2011) also find support for this argument. They argue that due to presence of a single consolidated tape, sophisticated order routing technologies and prohibition on trade-throughs, equity markets in the US are virtually consolidated allowing them to enjoy the benefits of consolidation even while remaining fragmented.
unlisted trading privileges. There was also the “third market” made up of non-member firms like Jefferies and Company. Before the advent of competitive commissions and abolition of NYSE Rule 394 (discussed in the introduction) these dealers competed by charging lower commissions. Later, they starting specializing in different segments of the market. For example, Jefferies and Company specialized in institutional order flow and Madoff Securities in retail order flow. Finally, alternative trading systems (ATS) such as Instinet and ITG’s POSIT, also competed for order flow by allowing market participants to anonymously trade with each other. The SEC, due to continued growth in the market share of such ATS and shortcomings in the existing regulatory framework, introduced order handling rules in 1997 and Reg-ATS in 1999 with the objective of integrating these systems into the NMS and encouraging effective competition among different venues.

In the EU, each member state had its own set of rules and regulations governing their national exchanges. In the 1980s and 1990s several initiatives were implemented in London, Paris and elsewhere with the objective of deregulating exchanges. London Stock Exchange’s Big Bang in October 1986 was by far the most comprehensive. By eliminating fixed commissions, allowing firms to operate in a dual capacity as brokers and dealers (or jobbers), restructuring the rules governing exchange memberships and adopting screen-based trading, London was able to capture order flow from several Continental European exchanges and it became the most dominant international center trading European securities. Simultaneous efforts to create a single market for financial services in the EU took the form of several directives like the admission directive of 1979, information directive of 1980 and the insider trading directive of 1989. However, the most significant step was the adoption of the investment services directive in 1993, which granted securities firms a passport to conduct cross-border operations in the EU. Additional rules focusing on concentration, off-market trading and price transparency were adopted with the objective of encouraging fair competition between exchanges and off-market venues.

In 2005, in one of the most significant overhauls of the NMS, the SEC introduced Reg-NMS with the objective of strengthening the NMS, increasing competition among individual markets and individual orders, and serving the interests of long-term investors and listed companies. In 2007, MiFID was operationalized in the EU in order to improve the efficiency of the single market.

---

8See Stoll (1992) for a summary of the different sources of competition in the 1980s in the US equity market.

9Clemons and Weber (1990) provide a summary of the changes implemented by the London Stock Exchange as part of the Big Bang and the impact they had on European financial markets.

to harmonize provision of trading services and to promote competition and consumer protection in investment services. As mentioned in the introduction, these changes led to a proliferation trading venues which competed with each other for order flow using innovative strategies like maker-taker pricing, low latency, and introduction of dark liquidity pools.\footnote{Hatheway et al. (2013) and Gomber and Pierron (2010) provide a classification of the different trading venues, post Reg-NMS in the US and post MiFID in the EU respectively.}

5. The Role of Trader Motives

Studies like Bagehot (1971), Kyle (1985) and Glosten and Milgrom (1985) examine the role of adverse selection in the price setting behavior of specialists or market-makers. These studies argue that, as specialists and market-makers have to trade with both informed and uninformed (or liquidity-motivated) traders, they charge a bid-ask spread in order to recover their losses from trading with informed traders. Further, by trading with informed traders, they learn about new information which is then incorporated into prices, making prices more efficient. In a competitive market with multiple venues, the informed traders’ choice of venue decides the level of contribution each venue make to the price discovery process. At the same time, the presence of uninformed traders allows specialists and other liquidity providers to profitably continue providing liquidity. Indeed, as observed in Glosten and Milgrom (1985), a high concentration of informed traders can lead to market failure as liquidity providers will be unwilling to trade at reasonable prices. In this section we review studies which focus on two issues while examining the role of informed and uninformed traders in a competitive marketplace: the cream-skimming hypothesis and trading in the upstairs and over-the-counter market. The former looks at the destabilizing role played by alternative venues and dealers by attracting uninformed traders away from exchanges, and the latter focuses on the positive aspects of uniformed trading in the upstairs and over-the-counter market.

5.1 Cream-Skimming Hypothesis

As discussed in the previous section, in response to the removal of barriers to competition by regulators, several dealers in the OTC market specialized in providing liquidity to retail traders. One of the practices that evolved as a result was that of dealers paying brokers for routing retail order flow (which was believed to be largely uniformed and hence profitable) to them and promising to execute such orders at the National Best Bid or Offer (NBBO). This practice attracted considerable controversy as it was argued that
the orders left over at the NYSE were large, informed and difficult to execute, leading to a deterioration in the quality of prices at the NYSE.

Easley et al. (1996) examine the impact of this practice on overall market quality by developing a continuous-time sequential trade model containing two venues with competitive risk-neutral market-makers providing liquidity. Informed and uninformed traders randomly arrive in the market and execute their orders on either of the two venues based on exogenous execution probabilities. Market-makers are informed about trades from both venues and they use this information to adjust their quotes using the Bayes rule based on the posterior probability of an information event. The authors argue that in the absence of alternative venues cream-skimming uninformed order flow from the primary listing venue, the probability of an informed trade and the fraction of total informed and uninformed trades on both venues should be equal. They test these hypotheses for trades on the NYSE and the Cincinnati Stock Exchange for 30 NYSE-listed stocks during 3 months of trading in 1990 and find evidence consistent with the cream-skimming argument. Specifically, the find that the information content of trades on the two venues is significantly different and the probability of an informed trade on the NYSE is 44% higher than on the Cincinnati Stock Exchange. The authors further argue that the resulting adverse selection problem is not only harmful for prices on the NYSE, but can also negatively affect prices on the cream-skimming venue, as prices on the NYSE are most of the time also the best prices in the market. The authors conclude that the profitability of such a strategy for the cream-skimming venue is directly related to its ability to successfully attract uninformed order flow and the overall welfare implications depend on whether and to what extent these gains are passed on to the retail customers.

Battalio (1997) examines the entry of Madoff Securities in NYSE-listed securities to examine whether it competes with the NYSE using a cost advantage (lower transaction costs) or an informational advantage (cream-skimming uninformed order flow). The author compares two measures of liquidity, time-weighted quoted bid-ask spreads and size-weighted effective spreads, before and after the firm enters the market, and finds that quoted spreads decrease across all firms and the effective spreads decrease for orders executed in one-eighth markets, suggesting that Madoff Securities competes with the NYSE using a cost advantage. Bessembinder and Kaufman (1997) compare execution costs for NYSE-listed securities on the NYSE and on NASDAQ and other regional exchanges. They compare quoted spreads, effective spreads and realized spreads and find evidence consistent with the cream-skimming hypothesis. While effective spreads show a modest increase for trades executed off-NYSE, the realized spreads (which measure
the revenue for market-makers) are two to three times larger for off-NYSE trades. More recent empirical studies like Degryse et al. (2013) and Hatheway et al. (2013) (discussed later) have also observed that dark venues successfully cream-skim uniformed order flow from the primary listing venues in US and European equities markets leading to deterioration in overall market quality.

One study which provides a positive view of market segmentation based on informed and uninformed order flow is Barclay et al. (2003). The authors study competition between NASDAQ market-makers and ECNs for 150 NASDAQ NMS stocks and one month of trading in 2000. They find that ECN trades are more likely during times when volume and price volatility is high. Using the Hasbrouck (1991a,b) vector autoregressive model, the authors also find that ECN trades have a 50% higher market impact than NASDAQ trades and they also contribute to about 2/3rds of the fundamental price variance. These results suggest that ECN trades are more informed. The authors further observe that NASDAQ trades are larger and uninformed as NASDAQ market-makers use internalization and preferencing agreements with brokers to successfully attract uninformed order flow. In conclusion, as opposed to Easley et al. (1996), Bessembinder and Kaufman (1997) and others who document cream-skimming of uninformed order flow away from the primary listing exchanges, the authors find that ECNs use anonymity and speed of execution to attract informed order flow and make large contributions to price discovery thereby enhancing overall market quality.

5.2 Trading in the Upstairs Market

The upstairs market, primarily used to trade large blocks over-the-counter, comprised of competing broker-dealers who either executed a customer order on their own account or searched for potential counterparties for their customers by shopping the block. Prices in this market were determined through bilateral negotiations. In this section, we look at two models which examine competition between off-exchange dealers and on-exchange specialists, focusing on the lack of anonymity in the upstairs market and how uninformed traders benefit from it (Seppi (1990)) and the role of unexpressed order flow (Grossman (1992)).

Seppi (1990) analyzes the existence and properties of the equilibria in the presence of information-based block trading. The author models the market for a risky security as a finite horizon multi-period market containing competitive specialists, off-exchange dealers, a group of small noise traders and a single large investor who may or may not

---

12 The nomenclature ‘upstairs’ market comes from the fact that these dealers used to be physically located above the floor of the NYSE.
be informed. The noise traders only trade in round lots with the specialist and the large
investor has the option of submitting a block trade to a dealer at the start of trading
or a sequence of anonymous orders to the specialist on the exchange. The author shows
that for a block with size less than a critical threshold, a separating equilibrium exists
in which the uninformed investor chooses to execute a block trade with the dealer but
the informed investor prefers to trade using market orders on-exchange. This is because,
as the dealer market does not provide anonymity, it allows the uninformed investor to
credibly signal to the dealer that she is uninformed by making commitments beyond
the terms of trade (for example, a no bagging commitment in which she agrees that she
will not trade the stock again for a few days). As opposed to the literature discussed
earlier where alternative venues attract uninformed or informed order flow away from
the exchange through appropriate incentive schemes or by adopting certain types of
technology, the author provides a framework in which informed and uninformed block
trades can be endogenously traded on different venues.

Grossman (1992) examines the role of the upstairs broker-dealer as a repository of
information about unexpressed order flow. The author models a two-period market
for a risky asset in which uninformed investors choose to trade with a specialist on
the exchange or in the upstairs market with a dealer. The specialist and the dealer
choose their demand and set prices in order to maximize an exponential utility function
of terminal wealth. An exogenously specified fraction of the uninformed order flow is
expressed in the first period and total order flow is exogenously distributed between
the two venues. Both the specialist and the dealer can observe the order flow directed
to their respective desks. However, as the dealer stays in contact with customers the
author assumes that she also observes the unexpressed order flow in the first period,
which is beneficial for her and detrimental to her customers. Additionally, customers
trading in the upstairs market are further disadvantaged because they have to search
for the best prices and have no assurance that the prices offered to them are the best.
On the other hand, in the downstairs market the specialist has less information about
unexpressed order flow, but investors do not have to search for the best prices. The
author concludes that, depending on these costs and benefits, either of the two market
structures will be superior in equilibrium and an interior equilibrium may also exist in
which both structures coexist, however such an equilibrium is unstable.

Both models have been extensively tested for trades in the upstairs markets all
over the world. Madhavan and Cheng (1997) analyze block executions in DJIA stocks
between December 1993 and January 1994 to empirically examine the role of upstairs
and downstairs market in providing liquidity for such trades. They use an endogenous
switching regression model to correct for endogeneity between trading costs and traders' choice of venue and find evidence consistent with Seppi (1990) that traders who can credibly signal that they are uninformed use the upstairs market. Smith et al. (2001) and Booth et al. (2002) find similar evidence for upstairs trading on the Toronto Stock Exchange and Helsinki Stock Exchange respectively. These studies find that the upstairs market provides substantial liquidity to large liquidity-motivated trades and plays a complementary role to the downstairs market without imposing costs on the overall market. Finally, these studies also find that large number of trades executed in the upstairs market offer price improvements, which is consistent with Grossman (1992) who argues that upstairs dealers are able to provide price improvements when they have information about customers' unexpressed demands. Additionally, Grammig et al. (2001) find similar evidence in the German stock market, where an anonymous electronic limit order book competed with the Frankfurt Stock Exchange, which operated a floor based trading system similar to the NYSE. They examine the relationship between trader anonymity and probability of informed trading using an augmented version of Easley et al. (1996) model, and find that the adverse selection component of the bid-ask spread and probability of informed trading are positively related, and are significantly lower on the Frankfurt Stock Exchange as compared to the anonymous limit order book.

Fong et al. (2001) examine the determinants of off-market trading, defined as upstairs trading, trading on crossing networks as well as after-hours trading, using data from the Australian Stock Exchange (ASX) between 1993 and 1998. Using a panel regression they find that off-market trading is primarily driven by institutional trading interest (trading volume and index inclusion), primary market liquidity (bid-ask spreads and market depth) and the existence of an options market. The authors also examine the difference in execution costs between the primary and upstairs market. Their results suggest that larger trades are more likely to be executed upstairs where they have a lower price impact, and an order of a given size is more likely to be traded upstairs if the primary venue is illiquid (higher spread and lower depth). Finally, the authors observe that there is an inverse relationship between price volatility and off-market trading. They interpret this as confirming the presence of a reputational signal in the upstairs market as suggested by Seppi (1990) as high price volatility makes it difficult to negotiate a price in the upstairs market and also to verify that a trader is uninformed.

While the upstairs market, as it existed in the 1980s and 1990s, does not exist anymore, these results are still relevant today as a large fraction of trading is generated in the over-the-counter markets which operate similar to the upstairs market. These markets are subject to less stringent regulation and are also less transparent. Many
OTC trading venues are in fact “dark” markets as opposed to regulated markets which are mostly “lit”. In Europe, the share of such OTC markets is considered to be around 40%, with regulated markets, multilateral trading facilities and systematic internalizers responsible for the rest. In Section 7 we examine the literature concerning dark markets.

In conclusion, it is important to recognize that new venues can compete with established exchanges on bases other than explicit costs. The studies discussed in this section on informed versus uninformed trading and on the upstairs market for large block transactions are cases in point. It is also important to recognize that not all types of competition are beneficial for the markets, and hence policy-makers need to be clear about what kind of competition they want to encourage.

6. Competition in Limit Order Markets

The studies reviewed so far have largely examined competition in specialist markets like the NYSE. However, most European markets like the Paris Euronext, London Stock Exchange and Deutsche Börse Xetra are organized as electronic limit order markets with or without the presence of designated market-makers providing liquidity. According to Jain (2003), 78% of world equity market capitalization is listed on such exchanges. Moreover, over the last two decades, even the role of specialists on the NYSE and other markets has been considerably diluted such that they also increasingly resemble limit order markets. In this section we review the research examining competition in limit order markets.

As noted by de Jong and Rindi (2009), models of specialist markets either operate a uniform pricing rule as in Kyle (1985), or have the same marginal price for each unit as in Glosten and Milgrom (1985). Limit order markets on the other hand operate a discriminatory pricing rule by allowing the same order to run up or down the limit order book and execute at multiple prices. Glosten (1994) derives the equilibrium in an electronic open limit order book where competitive risk-neutral traders provide liquidity to rational risk-averse investors (who may be informed) who choose their trade quantity to maximize a quasi-concave utility function of terminal wealth. The author shows that the bid and ask prices are equal to the lower-tail and upper-tail expectations of the value of the security, and such a limit order book mimics competition amongst anonymous exchanges. Due to their ability to average profits across informed and uninformed trades, limit order books achieve the best possible outcome in the presence of extreme adverse selection problems. In this sense an open limit order book is consid-

---

13 This ability of averaging profits across trades is one of the reasons why a monopoly in markets
ered competition-proof. Seppi (1997) analyses competition for liquidity provision in a hybrid specialist/limit order market (like the NYSE) by incorporating the timing difference between liquidity provision by specialists and other limit order traders; while limit orders are ex ante pre-commitments to provide liquidity to market orders arriving in the future, a specialist provides liquidity by providing ex post price improvement after a market order has arrived. The author finds that a hybrid market provides better liquidity to large institutional trades as well as to small trades, but a pure limit order market is beneficial for medium-size orders. Viswanathan and Wang (2002) compare investor welfare in three separate market structures, a pure limit order market, a dealerships market and a hybrid market, and find that risk averse investors always prefer a dealership market over a limit order market, however a properly structured hybrid market dominates a dealership market as well as a pure limit order market in terms of customer welfare. A second set of models by Parlour (1998), Foucault (1999) and others model limit order books as sequential games which allows them to endogenize the choice between limit orders and market orders. These models and their extensions along with the models discussed above have become the bedrock of the microstructure literature on limit order markets and have been extended to study competition between multiple limit order markets, or between a limit order market on one side and a specialist or dealer market on the other.

Parlour and Seppi (2003) extend Seppi (1997) by analyzing order flow competition between a pure limit order book, and a hybrid market containing a specialist and a limit order book, in a single period model. The authors show that, as opposed to Glosten (1994), if liquidity providers have heterogeneous costs, multiple equilibria can exist and neither the pure limit order market nor the hybrid market is competition-proof. In their model the order preferencing rules which determine where orders are routed when investors are indifferent determine market viability. Finally, the authors also find that the impact on aggregate liquidity is ambiguous. Foucault and Menkveld (2008) extend the Parlour and Seppi (2003) model to analyze competition for order flow in a market with multiple competing limit order books, focusing on the impact of trade-throughs on market quality. They examine the entry of the London Stock Exchange (EuroSETS) was considered valuable. As trading with uninformed investors allows specialists and market-makers to recover losses from trading with informed investors, they can stay open for business when other market structures shut down. This was one of the implications in Glosten and Milgrom (1985) as well as in Glosten (1989). However, Glosten (1994) finds that an open limit order book also shares this feature. The authors arrive at a different conclusion from Glosten (1994) and Seppi (1997) because they allow all market-makers to compete strategically as opposed to Glosten (1994) who examines monoplistically competitive limit order submitters and Seppi (1997) who allows strategic behavior on the part of specialists but not other limit order traders.
in the Dutch equity market and find that the consolidated order book is deeper after EuroSETS entry due to the absence of time priority, that is the ability of traders to queue-jump the incumbent venue by submitting limit orders in other markets. The authors also find that a large number of smart order routers (brokers who can route their orders to the market providing the best price) in the market increases liquidity supply on the EuroSETS due to an increase in execution probability. Based on their results they argue that protection against trade-throughs is important in fragmented markets.

Friederich and Payne (2007) analyze the determinants of on- and off-order book volume in the London equity market where a network of broker-dealer firms in the OTC market compete with a central electronic limit order book. Using an intraday panel estimation approach the authors find that on-order-book trading volume is low if execution risk, as proxied by total traded volume and average transaction size, is high, as well as when information risk, as proxied by trade imbalance, is high. This is because when execution risk and information risk is high, the liquidity service supplied by the dealers is valuable. Additionally, the authors also find that order flow routed to the limit order book is low when liquidity on the order book (bid-ask spreads and depth) is low. Other empirical studies like Hengelbrock and Theissen (2009), Chlistalla and Lutat (2011) and Gomber et al. (2011) have also observed a positive impact of competition between similarly organized limit order markets. Additionally, Boehmer and Boehmer (2003) document a similar positive impact of increased competition in the ETFs market after the NYSE started trading ETFs primarily listed and traded on the AMEX.

Based on the literature discussed above, it is clear the neither a pure limit order market nor a specialist market can be considered competition-proof. It is also clear that structural differences between limit order markets and specialist markets cannot be ignored while analyzing competition between different types of venues. While the presence of multiple competing limit order books in a market seems to be beneficial for market quality, it becomes difficult to arrive at a consensus the moment a limit order book is combined with another market structure like a hybrid market.

7. Market Transparency

O’Hara (1995) defines market transparency as the ability of market participants to observe the information in the trading process. Market transparency is important because the information available in the trading process can affect the strategies of market participants. Two issues are relevant here: what information is observable and to whom is
For example, what information should be made available to market participants in limit order markets? Should information about depth be made available along with limit order prices? How many levels of the order book should be disclosed? Additionally, should the information be disclosed to all market participants or should designated market-makers be given privileged access to certain information? We first examine markets with different levels of transparency and the migration of order flow to different venues due to these rules. Then we review the research which examines the impact of the relatively recent phenomenon of dark pools on market quality and overall welfare.

7.1 Disclosure/Transparency

Biais (1993) compares the strategies of liquidity providers in fragmented and consolidated markets by modeling the market for a risky security with risk-averse liquidity providers who compete with each other for a single market order from risk-averse liquidity demanders. The author assumes that, as opposed to consolidated markets where liquidity providers know their competitors’ inventories (through observed trades and quotes), in fragmented markets liquidity providers only know the distribution of competitor inventories. Liquidity suppliers’ quotes are determined in two steps. In the first step reservation quotes (quotes at which liquidity suppliers are indifferent between trading and not trading) are determined and in the second step the surplus maximizing optimal quotes at which trades can take place are determined. The central and puzzling result is that expected quotes and spreads under both market structures are identical. This is because in both cases the best quotes are set such that they are slightly better than the second best reservation price. Although risk aversion and differences in inventory positions affects liquidity supplier’s private valuations, due to the linear approximation used while calculating the surplus maximizing quotes, risk aversion does not influence the preferences of liquidity suppliers over the surplus from trade. However, bid-ask spreads are more volatile in fragmented markets. This is because, in fragmented markets, the best quote providing agent slightly improves her expectation of the second-best reservation price, whereas in consolidated markets, she exactly knows the second-best reservation price.

Madhavan (1995) analyzes the impact of disclosing trading information to market participants in a fragmented market. Using a modified version of the Glosten and Milgrom (1985) model, the author shows that if trade disclosure is voluntary, consolidation is not inevitable as predicted by Pagano (1989a) and others. The two-period model contains three types of traders: small liquidity traders who trade in a single period and
large liquidity and informed traders who break their order into two trades. As the large traders effectively front run their own trades leading to higher transaction costs, they prefer to trade in fragmented markets where trade disclosure is not mandatory. The author shows that non-disclosing dealers also prefer such markets as they can profit from the reduced competition and by observing the previous trade. Further, if some dealers are required to disclose their trades and others are not, the non-disclosing dealer can undercut the disclosing dealer in the first period and profit from the information in the later period. Finally, as less information is impounded into prices, fragmented markets have higher price volatility and are less efficient due to dealers paying traders for order flow.

Using an experimental approach, Bloomfield and O’Hara (2000) examine whether markets with different degrees of transparency survive in the presence of adverse selection. They develop a two-period sequential trade game-theoretic model containing dealers operating in two markets with differing levels of transparency. The dealers set bid and ask quotes at the start of each round. In each round an informed or liquidity motivated trader is randomly chosen. As the dealers operating in the less transparent market are able to profit from the information they learn by trading in the first period, they are able to quote narrower spreads and earn more profits. The authors test their model by conducting two laboratory experiments: In the first experiment, the traders operating in the low transparency market are exogenously specified and the experiment gives results consistent with their model. In the second experiment where traders are allowed to choose their transparency regime, the authors find that as the number of less transparent traders increases, their profits drop. However a small number of high transparency traders persist as they benefit from trades with informed traders who use them to broadcast deceptive trades.

Hendershott and Jones (2005) conduct an event study around Island ECN’s decision in September 2002 to go dark by no longer displaying its limit order book, and its decision to switch the lights back on in October 2003, for three actively traded ETFs. The authors analyze the impact of this decision on transaction costs and price discovery. The lack of transparency and the reduction in competition within Island leads to an increase in transaction costs (and adverse selection) on Island. On the other hand the venues that gain market share experience an increase in competition and a reduction in trading costs. On an overall basis however, Island’s decision negatively affects transaction costs. The authors also observe that around 50% of the trading volume remains within Island after its decision to go dark, thereby supporting the argument made by Harris (1993) that different markets serve traders with different preferences. After its decision to redisplay
the limit order book the authors observe that, as Island no longer dominates trading, the entire adverse effect of the earlier decision is not reversed. Madhavan et al. (2005), while studying the impact of an increase in pre-trade transparency on the Toronto Stock Exchange, find that market quality deteriorates after this change as limit order traders are more reluctant to provide a free option to other traders.

Overall the above studies demonstrate that the issue of market transparency is central to the consolidation versus fragmentation debate. Clearly dealers prefer not to disclose their trades to competing dealers in other markets as it allows them to profit from the information about the trade. However, as far as the impact of introducing a less transparent venue in a fragmented marketplace is concerned, the evidence so far seems mixed.

7.2 Dark Pools

SEC (2010) reports that 32 dark pools operating in the US had a market share of approximately 7.9% as of September 2009. While the market share of dark pools in the US is small, as per a report by Rosenblatt Securities,15 between 2002 and 2010 it is expected to grow by 40%. The same trend has also been observed in Europe.16 Due to the varied business models employed by dark pool operators, it is very difficult to define dark pools. For example, SEC (2010)’s definition of dark pools as “alternative trading systems that do not provide their best-priced orders for inclusion in the consolidated quotation data,” does not provide any information about how they operate. Mittal (2008) notes that dark pools have multiple characteristics based on their ownership structure, pricing, access, constituency and other factors. He develops the following taxonomy of US dark pools, which we use for the purpose of this article:

- Public Crossing Networks: These are the most traditional forms of dark pools, which do not contain proprietary order flow of the operator. These dark pools cross buy and sell orders as they arrive in the system (typically at the mid-quote). SEC (1998) defines them as a system “that allows participants to enter unpriced orders to buy and sell securities. Orders are crossed at specified times at a price derived from another market.” Examples include ITG POSIT and Instinet CBX.

- Internalization Pools: These are similar to public crossing networks except that these are designed to primarily internalize the operator’s order flow. Examples

16 See FT article by Philip Stafford dated October 17, 2013 “European dark pool equity trading jumps 45%.”
include Credit Suisse’s Crossfinder and Goldman Sachs’ Sigma X.

- Ping Destinations: These dark pools only accept immediate-or-cancel orders submitted by customers which interact with operators’ own flow. Examples include dark pools operated by high-frequency trading firms like Citadel and GETCO.

- Exchange-Based Pools: This category contains registered dark pools operated by exchanges (like Nasdaq Cross and NYSE Matchpoint) as well as dark liquidity associated with hidden and iceberg orders.

- Consortium-Based Pools: These are a hybrid between public crossing networks (typically owned by a consortium as opposed to a single operator) and internalization pools (operate as an organization separate from the individual operators). Examples include LEVEL started by Citi, Credit Suisse, Fidelity Brokerage, Lehman Brothers and Merrill Lynch, and BIDS.

Due to the potential impact of dark pools on the price discovery process, market participants and regulators have raised concerns over their increasing market share. For example, IOSCO (2011) notes that if orders that would otherwise be publicly disclosed go dark, dark pools may hamper the price discovery process.

Early studies on dark pools focused on public crossing networks. Hendershott and Mendelson (2000) analyze the trade-off between a dealer market (certainty of execution but positive spread) and a crossing network (reduced cost but uncertain execution) in a model that includes a random number of informed and liquidity traders, competitive dealers who post quotes such that their spreads include adverse selection costs, inventory holding costs and fixed operating costs, and a crossing network where traders directly trade with each other and orders are matched at the dealer’s mid-quote. In their model, traders choice of the venue depends on their private reservation values, transaction costs associated with each market (submission costs plus execution costs if an order is matched) and the probability of execution. The authors find that the introduction of a crossing network has two opposite externalities. Initially, as the crossing network gains market share, it induces a positive externality through improvements in overall liquidity benefiting all trades. Once it obtains a critical mass, an increase in orders by low liquidity preference traders induces a negative effect as it leads to a crowding out of other traders on the same side of the market. This happens due to the absence of price priority on the crossing network. Finally, the authors also observe that using the dealer market as a market of last resort leads to wider spreads and more efficient prices when information is short lived or when there is no asymmetric information. On
the other hand, if information is long-lived, the introduction of a crossing network has the opposite effect. Degryse et al. (2009) also examine competition between a dealer market and a crossing network focusing on different levels of transparency (the ability of traders to see past order flow from both markets). They find that these variations in the level of transparency lead to differences in the predictability with respect to order flow, and neither the introduction of a crossing network nor an increase in transparency is necessarily welfare-improving.

Conrad et al. (2003) use a proprietary dataset of executions between 1996 to 1998 on primary markets, ECNs and crossing networks and find that crossing networks are mainly used to trade NYSE-listed securities whereas ECNs mainly focus on NASDAQ stocks. This is because crossing systems need a good primary market price discovery mechanism and a sufficiently large pool of liquidity, both of which are present for NYSE-listed stocks. On the other hand, ECNs provide good price discovery for fragmented stocks with high bid-ask spreads on primary markets (also shown in Barclay et al. (2003)). The authors also show that execution costs on average are lower on crossing networks (ECNs) by 30bps (66bps) than those on primary markets. They argue that given best execution obligations, in equilibrium these differences should be zero. The authors find evidence that lower costs on ECNs are mainly due to a transition towards such an equilibrium and those on crossing networks are due to a downward bias in opportunity costs as the dataset used by them does not contain completely unfilled orders.

In other empirical studies, Gresse (2006), using crossing network data from Posit between July 2000 to June 2001 on the London Stock Exchange, finds that low bid-ask spreads are associated with high crossing network volume. The data allows the author to distinguish between crossing network trading by dealers and institutional investors. She finds that the above relation holds only for dealers trading in the crossing network. This suggests the presence of risk-sharing benefits for dealers that probably outweigh a possible but small cream-skimming effect resulting from institutions conducting uninformed trades on the crossing network. Ray (2010) examines the decision of a liquidity trader to send market orders to an exchange or use a crossing network. The author

---

17Best execution obligations mandate broker-dealers to provide the best possible terms of trade to their customers. In the past the phrase terms of trades generally applied to prices. While this is still the case in the US, since the operationalization of MiFID in 2007, the EU has materially departed from this position by including additional characteristics like likelihood and speed of execution in the definition. See Macey and O’Hara (1997) for the legal and economic aspects of the duty of best execution, and Petrella (2010) for the different approaches to best execution taken by the US and EU post the implementation of Reg-NMS and MiFID respectively.

18Naes and Ødegaard (2006) also find similar evidence using data containing the Norwegian Government Petroleum Fund’s trades in the US stock markets for the first six months in 1998.
tests his model’s implications using realized monthly crossing network volumes from Posit, Liquidnet, and Pipeline Trading between June 2005 to June 2006 for NASDAQ listed stocks, and finds that in the cross-section crossing network market share shows an inverted U-shaped relationship with market liquidity.

Buti et al. (2011a) analyze the interaction of a dark pool and a limit order market using a model that contains small and large traders with personal valuations for an asset around its common value. Their model predicts that the introduction of a dark pool next to a limit order market leads to a migration of order flow to the dark pool, however this comes part and parcel with an increase in overall trading volume. The authors also observe that dark pool market share is high when the limit order book is characterized by high depth and narrow inside spreads. Finally, the impact of introducing the dark pool on depth in the limit order market is negative, and the impact on spreads is inversely related to stock liquidity.

Two papers that examine adverse selection in dark pools are Ye (2012) and Zhu (2013). Ye (2012) analyzes the trading strategy of an informed trader when a security trades in a dark pool along side a dealer market. The author develops a rational expectations equilibrium that includes endogenously determined price and execution probability using a variation of Kyle (1985). The model implies that price discovery is reduced in the presence of the dark pool and price impact and the probability of non-execution in the dark pool are positively correlated because of the informed trader’s strategic order submission choice. The model further implies that a policy granting everyone the same access to participate in the dark pool, as suggested by the Securities and Exchange Commission in the US, is harmful to price discovery. Zhu (2013) examines the impact of competition between a lit exchange (with a competitive specialist providing liquidity) and a dark pool (which executes orders at the exchange mid-quote) on price discovery in a model featuring risk-neutral liquidity and informed traders. The author shows that the liquidity traders are attracted to the dark pool, whereas informed traders prefer to trade on the transparent exchange. This is because correlated informed orders have a higher non-execution risk on the dark pool than uncorrelated uninformed orders. This self-selection results in information-based trading being concentrated on the exchange which enhances price discovery. However, this also leads to an increase in adverse selection risk and wider spreads on the exchange. The welfare implications of introducing dark pools depend on this trade-off between the impact on price discovery and liquidity.

Buti et al. (2011b), using a data set of daily dark pool trading volume reported voluntarily by 11 dark pools from the United States in 2009, analyze the determinants of dark pool market share, and the effect of dark pool trading on market quality and price
efficiency. Their results show that dark pool trading focuses on liquid stocks, and NASDAQ stocks have higher dark pool market share than NYSE stocks after controlling for liquidity. In the time series, they find that dark pool market share is positively related to overall trading volume, inside depth, bid-ask spread, and inversely related to intraday stock price volatility, absolute returns, and relative order imbalance. Cross-sectionally, they find that market quality, as measured by bid-ask spreads, price impact and stock price volatility, improve with dark pool activity. However, their results with respect to price efficiency are less clear. Nimalendran and Ray (2013) analyze high-frequency transaction data provided by an anonymous crossing network that offers different ways of executing transactions. They find that for illiquid stocks, trades executed using algorithms appear informed, as inferred from larger bid-ask spreads and price impacts of exchange trades in the respective stocks for periods of up to two hours after the trade in the crossing network. Results differ for liquid stocks, for trades by the crossing network’s brokerage desk, and for large negotiated block trades. Finally, Weaver (2011) uses trade data reported through various Trade Reporting Facilities (TRFs) for AMEX, NYSE and NASDAQ listed stocks and the month of October 2010 to examine the impact of order flow internalization by broker-dealers on market quality, and finds that the degree of internalization is positively related to bid-ask spreads, price impact, and stock price volatility.

7.3 Lit Versus Dark Fragmentation

One of the important questions, which has been examined in recent empirical studies, is whether lit and dark fragmentation impacts market quality and welfare differently. O’Hara and Ye (2011) consider the effect of total fragmentation on market quality in the US equity markets using Rule 605 data for the first six months of 2008. The authors measure market quality using transaction costs measures (effective spreads, realized spreads and execution speed) and price efficiency measures (short-term price volatility and variance ratios). Without distinguishing between dark and lit markets, they find that transaction costs are lower and execution speed faster for stocks with more fragmented trading. Furthermore, while short-term volatility increases with fragmentation, so does price efficiency in the sense that future price changes are less predictable. Finally, the authors also observe that small cap stocks experience the most improvements in market

---

19 Under SEC Rule 605, market centers in the US are obligated to publicly disclose execution quality statistics on a monthly basis for all trades that satisfy the following criteria: orders must be held; limit price must be less than ten cents from the quote; order must be straight market or limit order; and the order must be for ten thousand shares or lower.
quality. They attribute all these improvements in market quality to the presence of a single consolidated tape, sophisticated order routing technologies and prohibition on trade-throughs.

Degryse et al. (2013), analyze transaction data comprising Dutch large- and mid-cap stocks between 2006 to 2009 and find that dark (lit) fragmentation lowers (increases) overall liquidity as measured by aggregate depth within specified price intervals around the mid-quote. They observe that dark venues cream-skim uninformed orders increasing the informativeness and price impact of trades on lit markets. However, they do distinguish between individual dark markets and they aggregate trading in all types of dark markets as well as the OTC market. Finally, the authors also observe that fragmentation is detrimental for primary market liquidity, suggesting that its benefits are not enjoyed by traders connected only to the primary exchange.

Hatheway et al. (2013), using a proprietary dataset from NASDAQ for 120 stocks and the first quarter of 2011, find that the different regulatory structure applicable to dark venues in the US (exemptions from fair access requirements and use of sub-penny pricing) allows them to cream-skim uninformed order flow from lit markets. As compared to lit venues, dark venues have 60% to 80% lower adverse selection risk and they execute about 20% of their trades at sub-penny price increments. This is accompanied by lower returns for liquidity providers on the lit markets and higher transaction costs on both lit and dark markets. The authors conclude that dark trading is detrimental to investors welfare because its benefits (better execution due to absence of informed order flow) are overshadowed by the detrimental effect on the benchmark prices in lit markets. The only exception to this are large trades on dark markets which have lower transaction costs and also improve price efficiency.

In conclusion, it is still unclear whether the introduction of dark pools in equity markets has been beneficial. It is also possible that all types of dark pool trading activity may not have a uniform impact on the markets, given the different types of market structures that are clubbed in its definition. The lack of quality data has been a severe impediment in examining this question. However, there seems to be a definite consensus that the distinction between lit and dark fragmentation is critical and cannot be ignored.

8. Tick Size and Maker-Taker Pricing

Before concluding, we briefly review the literature examining the role played by tick size and maker-taker pricing in the fragmentation of equity markets.
8.1 Tick Size

Tick Size is the minimum amount by which the price of a security can change. In the US the tick size for lit venues is fixed by law at 1 cent, whereas in the EU each venue is free to choose its optimal tick size. However, since June 2009, European venues have agreed to a harmonized tick regime, which was compiled by the Federation of European Securities Exchanges (FESE).\(^\text{20}\)

Buti et al. (2013) model competition between a public limit order book (PLB) and a lit or dark venue that allows sub-penny pricing. They extend the Parlour (1998) model by including two price levels on each side of the public limit order book and a finer pricing grid for the sub-penny venue (SPV). Their model contains a group of broker-dealers who choose between executing their customers orders on the PLB or SLV. Additionally, these broker-dealers are exclusively allowed to post limit orders on the SPV. The model also contains regular traders who access the liquidity on both venues through a smart order routing (SORs) technology, which allows them to submit their orders to the market offering the best price. If the SPV is lit the SORs can perfectly choose the venue offering the best execution, whereas if it is dark then the SORs can only infer the state of the SPV. The authors find that the SPV attracts both limit and market orders away from the PLB and the impact on market quality depends on the liquidity of the stock and stock price. Based on this, the authors conclude that these two variables should drive the choice of the optimal tick size.

Biais et al. (2010) examine competition for liquidity provision between NASDAQ and Island ECN for 74 NASDAQ stocks and 2 months in 2001, focusing on the role played by tick size. They find that NASDAQ spreads reduced dramatically after it transitioned from a 1/16th tick size to a decimal pricing grid. At the same time Island spreads also decreased suggesting that liquidity provision there was not competitive. The authors also find evidence that liquidity providers were earning oligopoly rents. Based on their observations, the authors conclude that theoretical models should not assume perfect competition among liquidity providers even if there is active competition.

8.2 Maker-Taker Pricing

Maker-taker pricing refers to an incentive scheme operated by some venues where market participants providing liquidity are given a rebate. Colliard and Foucault (2012) analyze

\(^{20}\text{However, in 2011 NYSE Euronext unilaterally decided to change its tick regime for Dutch and French blue chip stocks. See FT article by Jeremy Grant dated January 28, 2011 “Euronext sparks outrage with tick size reduction.”}\)
competition between a pure limit order book and a dealer market focusing on the role played by maker-taker fees. In their model, investors with different private valuations and different levels of patience sequentially arrive in the market and either choose to trade using a limit order or market order in the limit order market or in the dealer market. The investors also use the dealer market as a last resort, in case a limit order does not execute on the public limit order book. The limit order market operator charges a fixed fee per trade which is broken into maker’s fee and taker’s fee. The authors study the impact of this fee on expected welfare and find that while a reduction in total trading fees increases gains from trade for investors, it also affects the relative payoffs of limit orders as limit order traders submit orders with a lower execution probability. Based on their conclusions the authors recommend imposing a non-competitive floor on trading fees while allowing fee-based competition to flourish between multiple limit order markets. They also suggest that while empirically testing the impact of competition on bid-ask spreads, it is necessary to separately examine the impact on cum-fee spread and raw spread.

9. Conclusion

Financial markets in the US and Europe have evolved from a largely consolidated structure comprising primary listing venues, regional exchanges and the OTC market, to a highly competitive but fragmented structure comprising several exchanges and alternative trading systems like crossing networks, ECNs, dark pools and multiple broker-dealers offering internalization services. These changes have been fueled by an increase in computational capabilities, improvements in networking technologies and several regulatory initiatives like Reg-NMS and MiFID. In this article we have reviewed the theoretical and empirical literature, which examines the reasons why markets fragment as well as the resulting impact on different aspects on market quality.

The literature surveyed in this article provides several important lessons. First and foremost, the reasons underlying market fragmentation need to be examined keeping in mind two opposing economic forces acting simultaneously in securities markets. On the one hand, due to the presence of strong network externalities, financial markets have a tendency to consolidate in space and time. This is because traders want to trade when and where other traders are trading. Additionally there are also economies of scale in concentrating order flow as it allows exchanges to amortize fixed costs over more trades. On the other hand due to heterogeneity in traders’ preferences it is difficult to envisage a single market that can cater to all market participants. This includes differences in
trader motives, the sizes they trade as well as their patience levels.

Second, while examining the larger welfare implications of market fragmentation, it is important to recognize that embedded in securities markets is the market for trading services as well as the market for the traded securities. While the reduced explicit costs resulting from competition in the market for trading services can be beneficial to market participants, competition in the market for the traded securities can be harmful as the quality of prices may be adversely affected. The theoretical and empirical literature examining the impact of dark, non-transparent venues seems to be a case in point.

The literature discussed above also leaves several questions unanswered. The reasons why different stocks fragment differently are not yet clear. For example, as per LiquidMetrix (2010), in 2010 London Stock Exchange’s (LSE) market share in FTSE 100 stocks ranged from 50% to 80% whereas the market share for BATS ranged from 2% to 15%. Similarly for DAX stocks, Deutsche Börse Xetra had a market share of 37% for Siemens and 75% for Commerzbank in 2010. Empirical studies have also found that the market share of the OTC market is significant. In Europe it is around 40% for equities as per Gomber and Pierron (2010). The reasons for this are also not clear. Last, but not the least, as compared to equities, research examining the impact of fragmentation/competition in other asset classes is scant, and in some cases non-existent. This question is especially important in light of the MiFID II proposals in Europe, which plan to impose increased transparency requirements and regulatory oversight for a significant portion of OTC trading in all asset classes.

Finally, it is clear that the issues involved in the design of financial markets are extremely important in helping financial markets to achieve their core objectives. The complex, subtle and simultaneous interactions of the forces discussed in this article, makes the task of designing optimal policy responses and regulatory interventions extremely difficult. Policy-makers, regulators and exchanges should carefully evaluate the impact of their decisions keeping in mind these issues.
Appendix A.  Evolution of Equity Markets - Key Milestones

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1969</td>
<td>Instinet, the world’s first electronic trading system (later referred to as an Electronic Communications Network (ECN)), goes live. Later, it became part of the “fourth” market competing with NYSE by allowing non-members to trade NYSE-listed stocks.</td>
</tr>
<tr>
<td>1971</td>
<td>NASDAQ begins trading 2,500 over-the-counter (OTC) securities.</td>
</tr>
<tr>
<td>1975</td>
<td>As part of a directive from the US Congress to reduce barriers to competition, the SEC proposes establishing a National Market System (NMS) which would include the Composite Quotation System (CQS)(^a) and the Intermarket Trading System (ITS).(^b)</td>
</tr>
<tr>
<td>1980</td>
<td>SEC passes Rule 19c-3 to repeal NYSE Rule 394 (and AMEX Rule 5) which prohibited exchange members from trading listed securities off-exchange.</td>
</tr>
<tr>
<td>1993</td>
<td>European Union (EU) adopts the Investment Services Directive granting securities firms, licensed by respective home states, to conduct cross-border operations in the EU. Additionally, several rules like price transparency are proposed to promote competition.</td>
</tr>
<tr>
<td>1997</td>
<td>In response to the NASDAQ controversy, the SEC introduces order handling rules to increase competition among NASDAQ dealers and Alternative Trading Systems (ATS) like ECNs, as well as to improve transparency.</td>
</tr>
<tr>
<td>1999</td>
<td>SEC approves Regulation-ATS, giving ATS(^c) an option to register themselves either as national securities exchanges or as broker-dealers. Further rules were imposed on ATS(^c) registered as broker-dealers to better integrate them into the NMS.</td>
</tr>
<tr>
<td>2001</td>
<td>All US stock markets convert to a decimal system of quotation, leading to a decrease in bid-ask spreads.</td>
</tr>
<tr>
<td>2005</td>
<td>SEC Introduces Regulation National Market System (Reg-NMS) to with the objective of promoting efficient and fair price formation across securities markets.</td>
</tr>
<tr>
<td>2007</td>
<td>Markets in Financial Instruments Directive (MiFID), proposed a few years earlier to harmonize provision of trading services by promoting competition and consumer protection in investment services, becomes operational in the European Union.</td>
</tr>
<tr>
<td>2013</td>
<td>The Council of the European Union finalizes its version of the MiFID II proposals.</td>
</tr>
</tbody>
</table>

\(^a\) CQS compiles and distributes quotes to different ITS-trading venues
\(^b\) ITS connects all major equity execution venues and permits trading across venues
References


### Recent Issues

<table>
<thead>
<tr>
<th>No.</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>Nicole Branger, Patrick Grüning, Holger Kraft, Christoph Meinerding, Christian Schlag</td>
<td>Asset Pricing Under Uncertainty About Shock Propagation</td>
</tr>
<tr>
<td>33</td>
<td>Gabriele Camera, YiLi Chien</td>
<td>Two Monetary Models with Alternating Markets</td>
</tr>
<tr>
<td>32</td>
<td>Gabriele Camera, Alessandro Gioffré</td>
<td>Game-Theoretic Foundations of Monetary Equilibrium</td>
</tr>
<tr>
<td>31</td>
<td>Dirk Bursian, Sven Fürth</td>
<td>Trust Me! I am a European Central Banker</td>
</tr>
<tr>
<td>30</td>
<td>Dirk Bursian, Markus Roth</td>
<td>Optimal policy and Taylor rule cross-checking under parameter uncertainty</td>
</tr>
<tr>
<td>29</td>
<td>Iñaki Aldasoro, Ignazio Angeloni</td>
<td>Input-Output-Based Measures of Systemic Importance</td>
</tr>
<tr>
<td>28</td>
<td>Nicole Branger, Holger Kraft, Christoph Meinerding</td>
<td>Partial Information about Contagion Risk, Self-Exciting Processes and Portfolio Optimization</td>
</tr>
<tr>
<td>27</td>
<td>Tobias H. Tröger</td>
<td>The Single Supervisory Mechanism – Panacea or Quack Banking Regulation?</td>
</tr>
<tr>
<td>26</td>
<td>Michael J. Brennan, Holger Kraft</td>
<td>Financing Asset Growth</td>
</tr>
<tr>
<td>25</td>
<td>Holger Kraft, Alexander Schmidt</td>
<td>Systemic Risk in the Financial Sector: What Can We Learn from Option Markets?</td>
</tr>
<tr>
<td>24</td>
<td>Andrej Gill, Nikolai Visnjic</td>
<td>Performance Benefits of Tight Control</td>
</tr>
<tr>
<td>23</td>
<td>Andrej Gill, Nikolai Visnjic</td>
<td>Insight Private Equity</td>
</tr>
<tr>
<td>22</td>
<td>Dirk Bursian, Alfons J. Weichenrieder, Jochen Zimmer</td>
<td>Trust in Government and Fiscal Adjustments</td>
</tr>
<tr>
<td>21</td>
<td>Stefano Corradin, Reint Gropp, Harry Huizinga, Luc Laeven</td>
<td>Who Invests in Home Equity to Exempt Wealth from Bankruptcy?</td>
</tr>
</tbody>
</table>