

Internalization, investor protection and market quality

Bruno Biais (IDEI) & Didier Davydoff (IEM)

July 17th, 2002

Executive Summary

Thanks to the competition between market centers, European stock exchanges have been significantly modernized over the last 20 years. To sustain market quality in Europe, it is important that regulators continue to promote the competition to supply liquidity. The US experience in the 1990s, in particular the Nasdaq case, shows that liquidity supply cannot be deemed a priori to be competitive, that certain market structures are in fact conducive to anti-competitive behavior, and that regulatory intervention in this context can be very useful to restore market quality.

The financial scandals which occurred recently in the US (and to a somewhat lower extent in Europe) also show that collusion, lack of transparency and discretionary intermediation practices can reduce market quality very significantly, as well as destroy public confidence in the market. In this context, it is very important that European regulation sets high standards of disclosure and fair practices to maintain market integrity.

Against this backdrop, the present paper studies the economics of internalization, in order to shed light on what regulation is appropriate and thus contribute to the revision of the Investment Services Directive. Broadly defined, internalization refers to the execution of customers' trades by financial intermediaries, off the primary market. Internalization is a particularly important topic in Europe, because a large fraction of orders is likely to be, at least partially, captive, i.e., investors' ability to choose which intermediaries to route their orders to is limited.

To set the stage for our analysis we first present the microstructure of European and US stock markets, and liquidity supply and price discovery in this context. Both in the US and in Europe, over the last twenty years, the role of limit orders and order books has been enhanced, and a rich flow of information has been disseminated on limit order books. This has fostered the competition to supply liquidity and enhanced pre-trade transparency. Consequently, liquidity and market fairness have been improved. This is an important and necessary condition for the development of equity markets, and the resulting improvement in savings opportunities for households and funding opportunities for firms.

The first section also describes internalization practices in Europe and in the US and offers a typology of different forms of trade execution, emphasizing how they differ in terms of disclosure and discretion.

In the second section we discuss the economics of internalization. Our analysis is based on the stylized facts presented in the first section, on results from the academic literature, and on reasoning from first economic principles. We show that if it reduces pre trade transparency, internalization will i) hinder price discovery, and ii) reduce the extent to which orders can attract liquidity from the other side of the market. The latter would increase transactions costs for investors. We also show that, if it reduces the extent to which customer orders can be displayed as limit orders, internalization will weaken the competition to supply liquidity. This, again, would increase transactions costs. Thirdly, we show that, if it leaves large discretion to intermediaries, internalization could i) lead to violations of price priority, and thus reduce the incentives to supply liquidity by posting limit orders, ii) reduce the fairness of the market. The latter would not only increase transactions costs for small investors but also weaken investors' confidence in equity market and thus curb the development of the equity culture in Europe.

Building on the insights developed in the first two sections, Section 3 presents our policy proposals. We argue that intermediaries internalizing order flow should respect price priority, best execution, and pre trade transparency, and apply non discretionary order handling and trade execution rules. One important regulatory tool in this context would be to adapt to the European context the Order Handling Rule introduced in the US in 1997 by the SEC to restore competitive practices. We also discuss how and why these strict rules could be alleviated in the case of very large trades.

Outline:

I) The trading process and internalization in Europe and the United States

I.1) The microstructure of stock markets in Europe and in the United States

I.2) Trade execution and internalization in Europe and the United States

II) The economics of internalization

II.1) Internalization and the competition to supply liquidity

II.2) Internalization and price discovery

II.3) Internalization and market fairness

III) Policy Implications

III.1) Under what conditions can internalization increase or decrease the quality of the market?

III.2) What types of regulation do we propose?

Internalization, investor protection and market quality

Introduction

Thanks to the competition between market centers, European stock markets have been significantly modernized over the last 20 years. To sustain market quality in Europe, it is important that regulators continue to promote the competition to supply liquidity. The US experience in the 1990s, in particular the Nasdaq case, shows that liquidity supply cannot be deemed a priori to be competitive, that certain market structures are in fact conducive to anti-competitive behavior, and that regulatory intervention in this context can be very useful to restore market quality.

The financial scandals which occurred recently in the US (and to a somewhat lower extent in Europe) also show that collusion, lack of transparency and discretionary intermediation practices can reduce market quality very significantly, as well as destroy public confidence in the market.¹ In this context, it is very important that European regulation sets high standards of disclosure and fair practices to maintain market integrity.

Against this backdrop, the present paper studies the economics of internalization, in order to shed light on what regulation is appropriate and thus contribute to the revision of the Investment Services Directive. Broadly defined, internalization refers to the execution of customers' trades by financial intermediaries, off the primary market. Internalization is a particularly important topic in Europe, because a large fraction of orders is likely to be, at least partially, captive, i.e., investors' ability to choose which intermediaries to route their orders to is limited. Several factors explain why it is costly to switch from an intermediary to another. First, brokerage services are often bundled with other financial services because of economies of scope between such activities as the management of bank accounts, savings products, lending, and custody. This is particularly relevant in Europe, where universal banks dominate the market. Such

¹ While collusion was observed between auditors and firms, or research analysts and proprietary traders, as well as among market makers, the adverse consequences of discretionary practices were observed in the context of IPOs, as noted by academic studies (see e.g. Chen and Ritter, 1999) as well as SEC and DJ investigations. In all these cases, lack of disclosure was instrumental to enable the development of

economies of scope are at play, not only for retail investors, but also for institutions, to which global brokers provide a wide range of services. Furthermore, it could be costly or unattractive for foreign investors to sever their relation with large brokerage houses based in the same country as them and rely on the services of European intermediaries. Indeed, information asymmetries, cultural differences, and agency problems are likely to be greater when dealing with foreign brokers.

To set the stage for our analysis we first present the microstructure of European and US stock markets, and liquidity supply and price discovery in this context. This first section also describes internalization practices in Europe and in the US and offers a typology of different forms of trade execution, emphasizing how they differ in terms of disclosure and discretion.

In the second section we discuss the economics of internalization. Our analysis is based on the stylized facts presented on the first section, results from the academic literature, and reasoning from first economic principles. We show that pre-trade transparency, strict enforcement of best execution rules, and the ability for all investors to compete to supply liquidity are necessary for market quality. Based on this analysis we outline which conditions must be satisfied so that internalization does not reduce market quality spurs anti-competitive practices. We also analyze how discretionary order handling, associated with internalization, could reduce market fairness.

Building on the insights developed in the first two sections, Section 3 presents our policy proposals. We argue that intermediaries internalizing order flow should respect price priority, best execution, and pre trade transparency, and apply non discretionary order handling and trade execution rules. One important regulatory tool in this context would be to adapt to the European context the Order Handling Rule introduced in the US in 1997 by the SEC to restore competitive practices. We also discuss how and why these strict rules could be alleviated in the case of very large trades.

unfair practices. Biais, Bossaerts and Rochet (2002) study how explicit IPO pricing and allocation protocols better cope with agency problems than discretionary ones.

I) The trading process and internalization in Europe and the United States

Broadly defined, internalization refers to the execution of customers trades by financial intermediaries, off the primary market. In this section, we first describe how trades are executed and prices are formed on the major stock markets, in Europe and in the US.² Against this backdrop we then describe and categorize the different forms of internalization prevailing in these markets. This sets the stage for our economic analysis of internalization in the next section, by highlighting i) how stock markets fulfill their price discovery and trade facilitation mission and ii) the different aspects of internalization interacting with these missions.

I.1) The microstructure of stock markets in Europe and in the United States

The role of stock markets is to facilitate the matching of buying and selling interests, at prices efficiently reflecting available information relevant for the valuation of shares. To fulfill this mission markets must discover prices efficiently and offer liquidity to investors. How the European and US markets deal with these tasks is discussed next.

I.1.1) Liquidity

For the market to fulfill its mission satisfactorily it must be liquid, i.e., investors desiring to sell or buy shares must be able to do so, without triggering a large adverse change in price, i.e., without incurring large market impact. The simplest way to express the desire to trade a security is to place a market order, i.e., an order to buy or to sell a given number of shares, at the prevailing market conditions.³ Such orders convey the demand for liquidity. Liquidity is provided to them by agents posting prices at which they stand ready to trade. This service is offered in several different ways in the major stock markets.

² For readability this initial description is kept brief. A more in depth discussion of the European market model will be offered in a separate paper.

European stock markets:

European stock exchanges have been radically modernized over the last twenty years. They now operate as electronic market places, in continuous time, and disseminate a rich flow of market information to computers in trading rooms in real time. Liquidity supply in these markets is provided by limit orders, i.e., orders to buy (resp. sell) a given quantity of shares if the price is not above (resp.) below a given limit. The limit order book collects all the orders placed in the market. The aggregate quantity available for trade at each price in the book is referred to as the depth.

The London Stock Exchange SETS system, the Deutsche Börse Xetra system, the Scandinavian OM system as well as Euronext's NSC operate electronic limit order books.⁴ Limit orders can be placed by financial intermediaries as well as by investors. These orders are an important fraction of the order flow in these markets, even at the retail level. For example, at Euronext Paris, in May 2002, 88.0 % of the small orders (for less than 5000 Euros) were limit orders, while for larger orders the proportion was above 95%.

As Euronext, the London Stock Exchange, OM and the Deutsche Börse, the Tokyo, Toronto, Madrid and Milan stock exchanges (among others) also operate as electronic limit order books. Numerous academic studies have documented how liquidity is supplied in this context.⁵ Price and time priority are strictly enforced in these markets (except in the case of large trades, where special procedures are allowed for.) The theoretical analysis underscores the importance of best execution rules to foster the competition to supply liquidity. Placing limit orders can be costly, as it involves disclosing one's trading intentions, and running the risk of adverse execution in case of sudden changes in market pricing. Investors and traders may be willing to place limit orders in spite of these costs only if, in exchange, they can benefit from price priority. In turn, price priority can be enforced only if best execution rules are applied.

³ While we refer here as market orders to all orders expressed without a price limit, Euronext uses a more refined characterization.

⁴ While in France all trades are completed within the order book, in the UK only 75% of the trades (on average) are completed through the order book.

The New York Stock Exchange

On the NYSE, liquidity is jointly supplied by investors placing limit orders, and by the specialist, who manages the limit order book and trades on his own account.⁶ As in European markets, price priority is strictly enforced on the NYSE. While the microstructure of the European stock markets and of the Nasdaq have been radically changed over the last twenty years, that of the NYSE has remained fairly stable, except for some improvements in technology (for example electronic order routing) or in the price grid (now decimalized).

Unlike in the European markets, where all participants have equal access to the limit order book, in the NYSE, the specialist has a privileged position. For example, while in the European markets the five best bid and ask prices in the order book, and the associated depth, are electronically disseminated to all market participants, on the NYSE only the specialist can directly observe the book. However, the NYSE plans to enhance the transparency of its market by also disseminating electronically the five best quotes on each side of the book.

The Nasdaq

The 450 Nasdaq dealers play a key role in liquidity supply in this market.⁷ These dealers post quotes at which they stand ready to accommodate liquidity demands. These quotes are disseminated on Nasdaq screens:

- « Level I » screen are disseminated to brokers: They display the National Best Bid and Offer (NBBO) as well as the last trade
- « Level II » screens offer the same information as in Level I , and in addition the quotes of all the market makers. They are disseminated to large institutional investors.
- « Level III » screens are available to market makers: Along with similar information to Level II screens they offer the ability to post quotes.

⁵ See for example Biais, Hillion and Spatt (1995) for France, Griffiths et al (2000) for Toronto, and Lehman and Modest (1995) for Japan.

⁶ Interesting academic studies, discussing the structure of the NYSE, documenting the respective roles of the specialist and of limit orders, include: Corwin and Lipson (2000), Goldstein and Kavajecz (2000), Harris and Hasbrouck (1006), Hasbrouck and Sofianos (1993), Kavajecz (1999), Lee et al (1993), Petersen and Fialkowski (1994) and Ready (1999).

⁷ There is approximately 10 dealers per stock on average on Nasdaq.

Along with automatic execution at the quotes they have posted, Nasdaq dealers also offer telephone based trade negotiations.

Until 1994, only dealers could offer liquidity by quoting prices on the Nasdaq market. Consequently, competition was muted, and they were able to earn market power rents. This was initially shown by the academic studies of Christie and Schultz (1994), Christie, Harris and Schultz (1994) and then confirmed by following public debate and judicial investigations.

In response, to foster competition, the SEC introduced the Order Handling Rule (11Ac1). Essentially, this rule compelled the Nasdaq market to publicly display the limit orders placed by investors. It applies to all exchange specialists and to NASDAQ market makers receiving customer limit orders.⁸ According to the Order Handling Rule, market makers receiving limit orders from customers can choose between three options: i) executing the order immediately, ii) routing it to another intermediary or exchange displaying orders, iii) or displaying it (price and quantity) within 30 seconds.⁹

In practice Nasdaq dealers most often chose to route their customers' limit orders to intermediaries displaying them. This was made possible, and in fact very effective, by the creation and development of Electronic Communication Networks: e--brokers, relying on web based platforms, which collect limit and market orders, and match them or display them on internet based order books.¹⁰ In 2002 ECNs have been estimated to capture 39.3% of the dollar volume of Nasdaq trading. The major ECN, Instinet, was estimated to represent 12% of the trading volume on Nasdaq in February 2002 , while Island amounted to 9.6%, Redi Book to 6.5%, and Archipelago to 10.5%.

As a result of this reform, liquidity was greatly enhanced on the Nasdaq market, as documented by several empirical studies (see, e.g., Barclay et al, 1999). Furthermore, the transparent electronic limit order book operated by Island, and its fine pricing grid enabled public investors to efficiently undercut Nasdaq dealers, as shown empirically by Biais, Bisière and Spatt

⁸ Odd-lot orders, block size orders (at least 10.000 shares or 200.000 \$) and "all or none" orders are not in the scope of the rule.

⁹ The maximum delay is less stringent if the order is received during the opening or the reopening of the NASDAQ market, or at the commencement of an IPO.

¹⁰ ECNs for the order handling rule are: Island, Terra Nova, Bloomberg Tradebook System, Instinet, REDI system (Spear, Leeds and Kellogg)

(2002). This induced Nasdaq to improve its market structure, by opting for a tighter tick size (see Biais, Bisière and Spatt (2002).)

We have asked Nasdaq officials about the respective significance of pre-negotiated trades and trades resulting from posted quotes being hit. They have told us that, over the recent years the role of pre-trade negotiations has been reduced. Except for blocks, most trades result from quotes being hit and it is rare for a market maker to be called, except in the institutional trading setting. This is particularly true for small trades (orders of 1,000 shares or less), which are automatically routed to the best quote by the Small Order Execution System (SOES). Last year this system was extended to SuperSOES so that now virtually any size order can automatically execute against the quotes.

I.1.2) Price discovery

Price discovery refers to the process by which market pricing is progressively adjusted, to reflect buying and selling interest and information relevant for the valuation of the shares.

Price discovery at the opening of the market

A crucial point in time for price discovery is the opening of the market. At this stage investors and traders need to take into account events which occurred and orders which were collected since the last close. Both in Europe (on the Deutsche Börse, Euronext, the Milan and Madrid markets) and in the US (on the NYSE) opening prices are set within a uniform price auction. All the limit and market orders present in the book or collected since the last close are aggregated into a supply curve and a demand curve. The opening price is set to maximize trading volume. All orders executed at the opening trade at this uniform price. The matching of orders in such an auction is useful to aggregate heterogeneous information,¹¹ and reconcile differences in the willingness to hold the security across investors.¹²

¹¹ See Goldreich, 2002.

¹² See the theoretical analyses of Mark Satterthwaite, from Northwestern University.

Opening prices on Euronext are set after an extensive and transparent preopening period, during which indicative prices are quoted, so that investors can adjust their orders to market conditions, and supply liquidity on the buy or sell side of the market when it is demanded, and rewarded. The empirical analysis of Biais, Hillion and Spatt (1999) documents the significant information content of these indicative prices, and their convergence process towards the end of the preopening period. This highlights the important role of pre-trade transparency in the price discovery process.

The opening price auction on the NYSE illustrates the privileged position of the specialist. Indeed, before the opening price is set the specialist first observes the orders in the book, and then places its own orders. Consequently the monopolist is in the situation of a monopolist trading against the residual demand or supply curve resulting from the aggregation of the limit and market orders (see the theoretical analysis of Stoll and Whaley, 1990).

There is no opening auction on Nasdaq.

Price discovery during the trading day

After the opening, electronic limit order markets, such as Euronext, Xetra, SETS, or Island operate as continuous auctions, where investors can, at any point in time place or revise limit orders or hit previously placed limit orders with market orders. Price discovery in this context reflects the placement of orders reacting to market conditions, such as the current spread and depth, or the past trades, as well as the competition between several traders and investors. Correspondingly, information disseminated because of pre-trade transparency requirements (such as the current spread and depth) as well as post-trade transparency requirements (such as past trades) is a crucial ingredient of this price discovery process. When all trades stem from previously posted quotes being hit, pre-trade transparency is high. This enhances the price discovery process. This differs from a pure dealer market where trades would be prenegotiated, such as Nasdaq before 1995. Nasdaq has moved away from this situation, however, and now entertains a much higher level of pre trade transparency: as discussed above the role of pre negotiated trades has been reduced, while information dissemination on quotes and depth has been steadily enhanced.

Over the last few years, the electronic limit order books operated by ECNs (such as Island, Instinet and REDI) have been competing with the Nasdaq market makers, on price discovery as

well on liquidity supply. Huang (2002) documents that, for many new economy stocks, Island plays a major role in the price discovery process. ECNs have become the dominant market for the QQQ Echange Traded Fund. ECN order books and trades can be observed on their own sites. For example Island has a freely accessible and very transparent site, disseminating in real time full information on the dynamics of its order book (www.isl.com). The best ECN quotes also are displayed on Nasdaq screens.

Nasdaq will soon operate a new quotation system, Supermontage, aggregating quotes posted by market makers and ECNs. While currently Level II Nasdaq screens display only the best bid and ask quotes of each participant, supermontage will display the aggregate interest in a stock five price levels deep on each side of the market. This could bring the Nasdaq market further to the European electronic limit order books, where over the last ten years, depth at the five best prices on each side of the quotes have been disclosed. This could make it easier for brokers to pick the best price and thus offer best execution to their customers. In turn, this could foster the competition to supply liquidity. While, always imposing price priority, Super Montage will allow for different priority regimes to be used in case of price ties. Time priority, taking into account access fees, size priority and preferencing are among the possible options.

Critics of this new system, however, emphasize that it will be costly, especially for ECNs. In particular, the Nasdaq fee formula will be based on the number of bids and offers posted each day, which is especially large on ECNs such as Island (where posting limit orders is free). This could undermine the ability of these ECNs to offer cheap execution, and possibly consequently deteriorate the supply of liquidity on this market.

I.1.3) Summary and conclusion

Both in the US and in Europe, over the last twenty years, the role of limit orders and order books has been enhanced, and a rich flow of information has been disseminated on limit order books. This has fostered the competition to supply liquidity and enhanced pre-trade transparency. Consequently, liquidity and market fairness have been improved. This is an important and necessary condition for the development of equity markets, and the resulting improvement in saving opportunities for households and funding opportunities for firms.

I.2) Trade execution and internalization in Europe and the United States

I.2.1) A typology

Several different forms of trade execution can take place within the market structures described above. These vary from transactions carried within the context of the main regulated exchanges (most often on limit order books), to trades within Alternative Trading Systems (such as ECNs or Crossing Systems), and internalization. Figure 1 offers a categorization of these different forms of trade execution. This classification of trade executions is related to the level of discretion enjoyed by the intermediary. While executions conducted within limit order books or crossing systems with explicit algorithms (depicted in the upper part of the figure) leave little discretion to financial intermediaries collecting investors' orders, internalization (which is depicted at the bottom of the figure) allows for greater discretion.

Orders posted and executed on an order book

The top of the figure refers to trades executed within a limit order book, either on the main exchanges, (e.g. Xetra (Deutsche Börse), NSC (Euronext), the NYSE) or through ECNs (e.g. Island). From a regulatory point of view, ECNs often have a broker status. Economically, however, they are more similar to exchanges than to intermediaries internalizing customer orders. As discussed above, they operate a full fledged transparent limit order book, competing with the liquidity and price discovery services offered by Nasdaq market makers. Furthermore, some ECNs choose to become exchanges. In line with the European Commission, we consider that these trades should be subject to regulated markets or Automated Trading Systems (ATSS) rules, rather than the internalization regulation.

On the Stockholm Stock Exchange's limit order book, orders placed by an intermediary can be executed in priority against other orders placed by the same intermediary. Thus, while price

priority must be complied with, time priority can be violated. In some sense this enables a form of internalization within the limit order book.

Orders not executed on a limit order book but executed within the context of a predefined trading algorithm

Such orders can be executed within crossing mechanisms which do not take proprietary positions, and merely match orders at predefined points in time, at prices previously set on the primary market. Posit offers an example of this type of mechanism.

Alternatively, orders can be routed to financial intermediaries which can execute them against their own inventory. For example, brokers operating on regional stock exchanges in the US can route their customers' orders to the local specialist, operating in the same firm as them, instead of sending these orders to the NYSE. Battalio, Greene et Jennings (1997) find that this type of internalization on the Cincinnati Stock Exchange amounted to 7% of NYSE orders in September 1994. Another example is offered by Madoff Securities, which matches retail orders against each other as well as against their own inventory. In both cases, the execution algorithm is predefined. Madoff securities matches orders at predefined points in time at or within the NYSE best quotes; brokers and specialists on regional exchanges execute orders, when they receive them, at or within the NYSE best quotes.

Orders not executed on a limit order book and for which the intermediary enjoys some discretion regarding the execution protocol

Financial intermediaries can enjoy some discretion relative to routing of the orders they collect or the timing of their execution. They can have the option to offer execution against their own inventory as soon as the orders are received (at or within the quotes currently prevailing on the primary exchange). Alternatively they can decide to delay execution somewhat, and later possibly pool customer orders received over a period or time. If there is an imbalance between the buying

and selling interest built during this batching period, the financial institution can decide to execute it against its own inventory. Alternatively, it could decide to route it to the primary market.

In the UK, Retail Service Providers specialize in the execution of the retail trades. 97% of the retail business on Blue Chips is handled by RSPs, who collect these orders from retail brokers and report their execution to the LSE, within the 3 minutes regulatory delay. FSA regulation 5-39 requires firms to provide best execution of their orders to private customers. The RSP is in charge of ensuring that the price is the best available -for transactions of the kind and size concerned- to the customer in the relevant market at that time. The rule book mentions that “in determining whether a firm has taken reasonable care to ascertain the price which is the best available for the customer in the relevant market, FSA will have regard to the conventions of that market”.

In Germany, financial intermediaries can match orders away from regulated markets. According to some professionals, such trades could amount to approximately 25% of Euro trading volumes in this country.

I.2.2) Internalization, transparency and best execution

Internalization and best execution

Internalized executions are not conducted without regard to market conditions on primary exchanges.

First, in most cases, internalized trades have to be reported to and put through a regulated exchange. For example, “applications” conducted on Euronext have to be immediately reported to the exchange, which includes them in its post-trade transparency flow. Another example is offered by Island. Since this ECN is not a regulated market it has to put its trades through an exchange, after executing them on its book. Initially most of these trades were sent to Nasdaq. However, from February 2002 to June 2002, Island has been completing the bulk of its trades on the Cincinnati Stock Exchange¹³, where fees are lower than on Nasdaq.

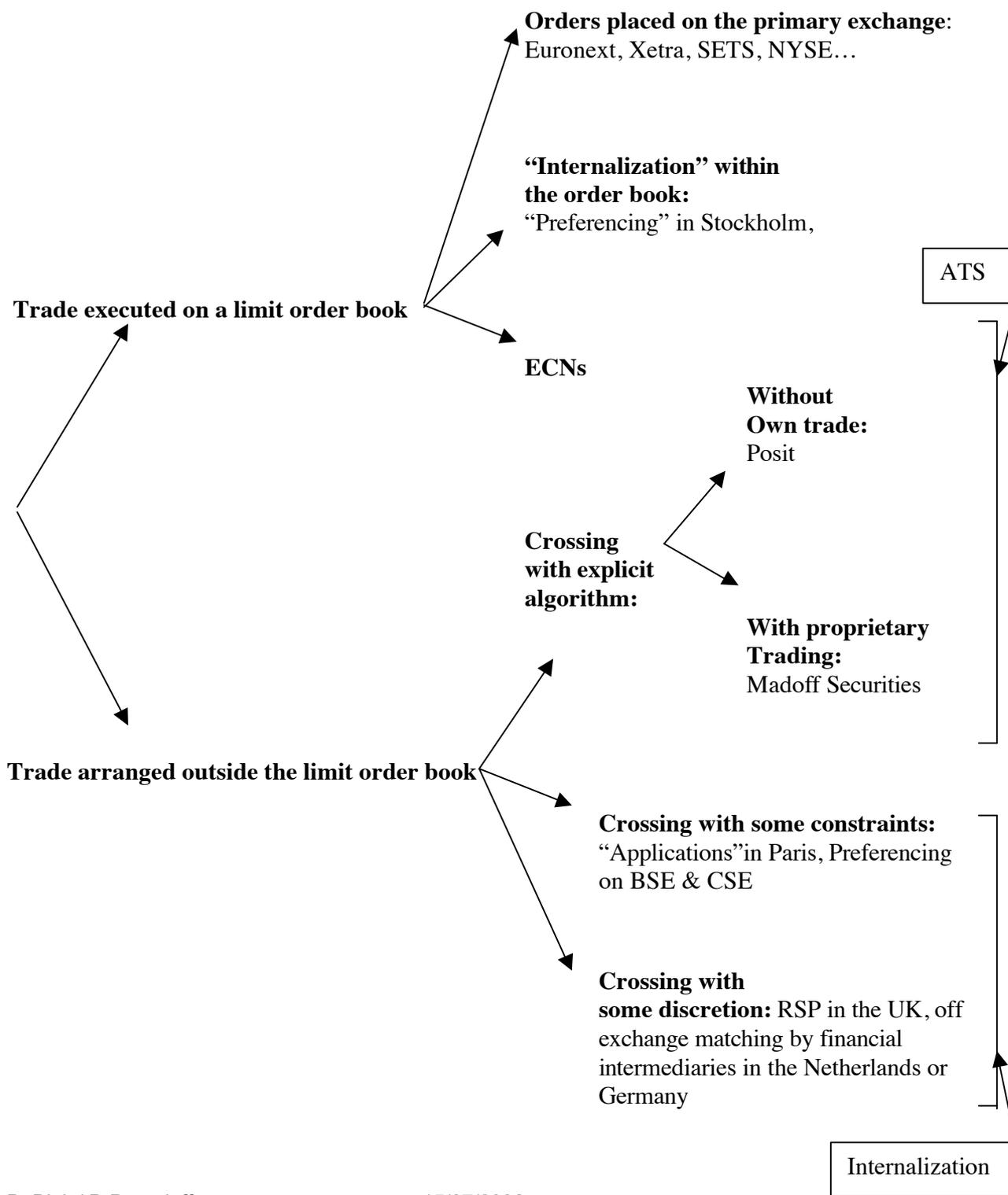
¹³ Although the price discovery and the operational matching of orders are processed by Island.

Second, financial intermediaries must achieve best execution for their customers. The SEC monitors and enforces brokers' compliance with this rule. In the recent past, however, regulators have criticized internalization, raising the concern that some brokers use their discretion to follow routing practices inconsistent with best execution. In November 1999, the SEC Chairman stated: "I worry that best execution may be compromised by payment for order flow, internalization and certain other practices that can present conflicts between the interests of brokers and their customers." A report issued by the General Accounting Office, the investigative arm of Congress, in June 2000, reiterated this warning. GAO investigators examined trading practices at 12 broker-dealers, with a significant market share. Six of those firms were sent a deficiency letter by the SEC accusing them of not ensuring that their customers' trades received the best execution possible. The report stated that "most of them routed orders to market centers whose execution quality ... were well below average."

Transparency

While the different forms of trade execution depicted in Figure 1 differ regarding the discretion enjoyed by the intermediary, they also differ in terms of transparency. In all cases, post trade transparency can be achieved, if requested by the regulatory authority. The level of pre-trade transparency achieved when trading outside the primary market is conducted within a limit order book can be as good as that achieved on the primary market, as shown by the example of Island. Crossing mechanisms often fail to offer pre-trade transparency. As long as they operate within predefined trading algorithms, however, they could be structured so as to achieve a very high level of pre-trade transparency. For example, indications on the current level of buying and selling interest could be disseminated before the crossing. This would contribute to the price discovery process.

Figure 1: Typology of different forms of trade execution



II) What are the consequences of internalization?

II.1) Internalization and the competition to supply liquidity

II.1.1) Market quality requires that liquidity supply be competitive

In the Wealth of Nations Adam Smith underscores that efficient markets outcomes can arise only if competition prevails. As emphasized by Adam Smith, this implies that many suppliers be able to offer their goods or services on the market, which, in turn, requires that they have enough information about market conditions. This reasoning directly applies to liquidity services. Tight spreads and large depth can obtain only if liquidity supply is competitive. This requires that many investors and traders be able to offer liquidity by posting quotes. This in turn implies that they have good information about market conditions, i.e., that pre- and post-trade transparency be high.

The Nasdaq case illustrates forcefully the lack of liquidity which is bound to result when the number of liquidity suppliers is limited and pre-trade transparency is limited.¹⁴ In that case, insufficient competition resulted from the fact that investor's orders could not be expressed as limit orders, competing with dealer quotes. Rather, they were deemed to be market orders, and could only hit dealer quotes. This is in stark contrast with a limit order book market where all orders, including those placed by small investors, can be displayed as limit orders, and compete to supply liquidity. This situation prevails on Island, or on the European stock exchanges: SETS, Euronext and XETRA.¹⁵

As discussed above, before 1995, Nasdaq dealers earned rents by quoting excessively large spreads, because public investors were not allowed to compete with dealers to supply liquidity by placing limit orders. In response, the SEC introduced the order handling rule to enable investors to place limit orders and consequently liquidity was vastly enhanced on the market (see Barclay et al, 1999).

¹⁴ While such non competitive liquidity supply has been evidenced empirically by Christie and Schultz (1994) and Christie, Harris and Schultz (1994), it has also been shown to be expected theoretically (see e.g. Dutta and Madhavan, 1997, and Biais, Martimort and Rochet, 2000).

¹⁵ As mentioned above, even small orders are often in majority expressed as limit orders. For example, on Euronext Paris, 88% of the small orders were limit orders in May 2002.

II.1.2) When the main market has monopolistic or oligopolistic features, crossing networks or internalization will arise to compete rents

As discussed above, in some respects, the NYSE specialist enjoys monopoly power, which is likely to create rents, and inflate transactions costs. As discussed in the previous section, several crossing systems and internalization practices have diverted order flow away from the NYSE. This is natural as such order flow diversion enables the intermediary to avoid passing rents to the specialist. Whether this results in gains for final investors is less obvious. Indeed, if intermediaries match the uncompetitive price demanded by the specialist for liquidity services, rents are simply transferred, not competed away.

Several empirical studies actually suggest that competition for order flow between the NYSE and crossing systems and internalizing intermediaries actually resulted in some improvements in liquidity. Battalio, Greene and Jennings (1997) study the impact of a reform which allowed brokers to execute their customer orders themselves on the Boston and Cincinnati Stock Exchanges without respecting the time priority of other dealers on these exchanges. They find that the ability of brokers to preference their own specialist units led to a substantial diversion of executions from the NYSE to these regional markets. As this took place, the NYSE spread actually declined. Similarly, Battalio (1997) finds that NYSE spreads are reduced after Madoff Securities begins purchasing order flow to attract order flow away from the NYSE.

The above analysis emphasize lack of competition between agents posting prices. Another problem is when markets are in such a position that they can charge non competitive fees. In presence of such inflated exchange fees, intermediaries have all incentives to set up or use alternative market places offering more economical conditions. This, in turn, can help incentivizing the major market to maintain low fees. To the extent that the fixed costs of turning to an alternative market place is very low, the logic of contestable markets can apply here, to maintain the fees of the main exchange close to their competitive level.

II.1.3) When the main market structure is competitive, internalization can severely reduce market quality

Internalization can reduce the quality of the market by fragmenting it

Now, assume alternatively that the structure of the main market is competitive. In this context consider the case where large brokers or financial intermediaries, collecting orders from investors do not input them in transparent limit order books, nor expose them to the public. Suppose instead they execute these orders internally, by crossing them against each other or against their own inventory, at or within the best quotes established on the main market.

This would reduce the pool of orders aggregated in the main market. As shown by the theoretical market microstructure literature, this, in turn, would reduce the liquidity of the main market, and, in particular, widen its spread.¹⁶ The economic intuition for this phenomenon is that liquidity is a positive externality: the more the other investors trade a given stock, the greater the liquidity of this stock, the lower the spread I pay when I trade it.

In this context, internalized orders trading at the best market quotes would incur larger transactions costs than if there was no internalization.

To better illustrate this example it may be useful to consider the following numerical example:

- First consider the case where all orders are routed to a competitive limit order book. Suppose the bid for stock XYZ in this situation is 100 Euros, while the ask price is 100,5 Euros. In this context, market orders are executed, for sure, at the cost of a 50 bp spread. Investors can also choose to place limit orders. While these are not executed for sure, they can obtain better execution. For example, consider a limit order to sell, placed at the previous mid-quote, say 100,25 Euros. This new order, at a relatively low price may attract additional liquidity from the other side of the market, if investors interested in buying felt that 100.5 Euros was too expensive, but 100.25 Euros is acceptable. Empirical analyses of order book dynamics offer evidence consistent with this reasoning.¹⁷

¹⁶ See the theoretical analyses of Pagano (1989) and Admati and Pfleiderer (1988).

¹⁷ See Biais, Hillion and Spatt (1995) for evidence of this phenomenon in the Euronext market structure. Griffiths, Smith, Turnbull, and White (2000) discuss interesting evidence from the Toronto Stock Exchange, in line with the Biais, Hillion and Spatt (1995) analysis.

- Now turn to the case where a significant fraction of the order flow is internalized, as described above. As discussed above, since this reduces the liquidity of the main market, it widens its spread, say by 20 bp, so that the bid price is now 99.90 Euros, while the ask is at 100,6 Euros. In this context, internalized market orders trading at the best quotes of the main market pay a 70 bp spread. What about internalized limit orders? Consider the example offered above, where a limit order to sell was placed at 100.25 Euros. This order could obtain execution if matched against the order of another customer.

This numerical example illustrates how internalization can increase the transactions cost incurred by market orders. It also shows that limit orders can fare better in this context, as suggested by the example of the order to sell at 100.25 Euros which could have obtained execution under the same terms in the two cases. Note however that the likelihood of execution of this order (and the associated gains from trade) can be expected to be lower in the internalization scenario:

- First, because this order has not been exposed, it cannot attract liquidity from the other side of the market. This point is in line with the conclusion reached by the US General Accounting Office report issued in June 2000 which stated that when orders are internalized, “other market participants, such as market makers or specialists, do not see them and, thus, do not have a chance to better the NBBO price.”
- Second, because it is not handled within an explicit algorithm, its price priority can be violated.

Internalization can deteriorate market quality market by reducing liquidity supply

By diverting limit orders away from the main market, internalization reduces the competitive pressure on those liquidity providers who still operate on the main market. This is likely to reduce liquidity and widen spreads on that market. In the limit, as the number of liquidity suppliers remaining on the main market becomes really low, this can create collusion problems,

similar to those which arose on the Nasdaq market up to 1994. Note that, in this context, widening the spread on the main market is costly for market orders addressed to that market, as well as to internalized orders if they are executed at the bid and ask quotes set on the main market.

As discussed above, to the extent that internalized orders are handled discretionarily by intermediaries, without pre-trade transparency and explicit algorithms, price priority may not be enforced. As discussed above, price priority is important to incentivize traders and investors to supply liquidity by placing limit orders. Consequently, investors who anticipate that their orders will be internalized can be discouraged from supplying liquidity supply. This is likely to reduce market quality.

An even worse situation arises if intermediaries internalizing orders do not offer investors the possibility to place limit orders, treating all orders as internalized market orders. In that case, to the extent that order flow is captive, market quality is very significantly reduced by internalization.

Auguy (2002) offers some insights on this point, using data from Euronext. For five of the most liquid shares, Auguy (2002) studies the consequences of removing from the order book customer limit orders, smaller than the average order size, transmitted by 9 large banks, members of the exchange. He finds that this very significantly reduces market quality. For example spreads widen by 13% to 35%.

II.2) Internalization and price discovery

II.2.1) If internalization reduces pre-trade transparency, it will impair the price discovery process

As discussed in the previous section, price discovery refers to the evolution of market pricing progressively incorporating relevant information on share valuations, supply and demand. As discussed above, pre-trade transparency facilitates this process by informing market participants of buying and selling interests, enabling investors to react and correspondingly prices to adjust. This is very much in line with the tâtonnement process studied by Walras in the *Eléments d'Economie Politique Pure* (1874). For example, Walras (1874) wrote (p. 48) that on the path to equilibrium:

“no exchange takes place ... without the sellers being able to lower the price and the buyers to raise it. ”

As discussed in the previous section, this process arises during the preopening period. It also arises during the continuous market session in limit order markets. For example, investors can express their desire to trade, and attract liquidity from the other side of the market by placing aggressive limit orders, for large quantities, within the spread or at the best quotes.¹⁸ Another example is offered by the workings of the limit order book on the Tokyo stock exchange. In certain circumstances, before executing incoming market orders, the market system issues a warning, alerting market participants to the imbalance in the order flow, and inviting them to step in to supply liquidity (see Modest and Lehman, 1995).

These market situations illustrate the dynamic interaction between orders demanding liquidity and orders supplying it. They highlight the crucial role played by pre-trade transparency, the disclosure of buying or selling interest, expressed as market or limit orders. They emphasize that price discovery is achieved via the reaction of market participants placing new orders in response to these indications of interest.

II.2.2) Exposing small and medium sized orders is important for price discovery

It should be emphasized at this point that the desirability and feasibility of order disclosure is not limited to large orders. Quite the contrary, it relates to all types of orders:

- As illustrated by the market mechanisms described above, disclosing market orders is desirable since these orders reflect buying and selling interest which should be taken into account to reach equilibrium pricing. It's also feasible, as illustrated by market mechanisms such as the preopening period in Euronext or the use of warning quotes on the Tokyo stock exchange.
- Small and medium sized orders are likely to have significant information content, as well as large orders. The information content of relatively small orders can result from order splitting, as shown theoretically by Kyle (1985). That medium sized orders are the most informative has been established

¹⁸ See the empirically analysis of Biais, Hillion and Spatt (1995).

empirically for the US market by Barclay and Wagner (1993) and for the UK market by Hansch et al (1999).

III.3) Fairness of the market

Public confidence is a necessary condition for markets to attract savings flows. Therefore one of the goals of the regulation is to ensure the fairness of the market. Registered financial firms have a fiduciary duty towards their customers. In particular they must provide best execution.

Discretionary order flow internalization can be at odds with this fiduciary duty. Indeed, conflicts of interest can arise in this context.

For example, financial intermediaries could time the execution of retail orders to fit their own proprietary interests or those of favored large customers.

Furthermore, as discussed above, exposing retail order flow in the public limit order book would attract competitive liquidity supply, while executing it internally would leave the financial intermediary in a monopolistic situation as regards liquidity supply. Consequently, investors end up paying a larger spread if their orders are internalized. Intermediaries, in contrast, earn larger rents with internalization.

Discretionary internalization, without pre-trade transparency can even create the risk of front running.

Consequently, discretionary internalization, without explicit order execution algorithms, and without pre-trade transparency, creates the risk of reducing the fairness of the market, thus lowering execution quality and discouraging investors from holding equity portfolios.

III) Policy Implications

The analysis offered in this section is based on the premises that the goal of financial markets regulation should be i) to avoid anti-competitive practices, in particular in liquidity supply, ii) to maintain the confidence in the markets, in particular by protecting investors' rights, and, more generally, iii) to maintain the quality of the market, in terms of liquidity and price discovery.

III.1) Under what conditions can internalization increase or decrease the quality of European stock markets?

As discussed in Section 1, European stock markets are operated as transparent, easily accessible limit order books. This market structure is quite competitive, as shown theoretically by Glosten (1994). In this context, Section 2 has established that internalization can weaken the competition to supply liquidity, impair the price discovery process, and deteriorate the fairness of the market. In particular:

If it reduces pre trade transparency, internalization will i) hinder price discovery, and ii) reduce the extent to which orders can attract liquidity from the other side of the market. The latter would increase transactions costs for investors.

If it reduces the extent to which customer orders can be displayed as limit orders, internalization will weaken the competition to supply liquidity. This, again, would increase transactions costs.

If it leaves large discretion to intermediaries, internalization could i) lead to violations of price priority, and thus reduce the incentives to supply liquidity by posting limit orders, ii) reduce the fairness of the market. The latter would not only increase transactions costs for small investors but also weaken investors' confidence in equity market and thus curb the development of the equity culture in Europe.

On the contrary, to the extent that it maintains high pre-trade transparency, that it enables customers' limit orders to be displayed, and that it is based on explicit, non-discretionary, trade execution protocols, internalization can enhance market quality. Indeed, it can enhance the pressure on regulated markets to keep exchange fees close to their competitive level.

III.2) What types of regulation do we propose?

III.2.1) Order handling rules, similar to those introduced by the SEC, are necessary for liquidity supply to be competitive

While equity markets in the US have progressively evolved towards enhanced pre-trade transparency, and better opportunities for investors to place limit orders, it is questionable why European markets should follow the opposite path. Rather we suggest that financial markets regulation in Europe should borrow from its US counterpart.

Indeed, it would be timely to introduce in Europe a regulation similar to the SEC Order Handling Rule (11Ac1). As mentioned above, this rule implies that intermediaries must either display their customers' limit orders or transmit them to exchanges or to systems, such as ECNs, offering limit orders display facilities. This would maintain the right for investors to compete to supply liquidity, and thus would contribute to avoiding anti-competitive practices. By offering pre-trade transparency to customer's limit orders, it would allow them to attract liquidity, as well as enhance price discovery.

III.2.2) Explicit, non-discretionary trade execution protocols are necessary to protect investors

Discretionary order handling raises the possibility that certain investors would receive unfair treatment. For example, privileged, large customers could be given better treatment than small investors. To avoid such unfair market situations, internalization should rely on explicit, non-discretionary trade execution protocols. These could include for example:

- Strict enforcement of best execution rules, and in particular ruling out execution of internalized trades away from the best quoted spread, computed taking into account prices posted on all trading venues (regulated markets, ATSS or other internalizing firms).

- Enforcing clear priority rules for internalized orders. The first of these rules should definitely be price priority. Time or size priority could also be used.
- Matching customer's orders at predefined points in time, as in crossing systems such as POSIT.

III.2.3) Strict enforcement of the best execution duty is necessary to protect investors and ensure that liquidity supply is competitive

Enforcing best execution rules is necessary to protect investors when multiple trading venues coexist. Foucault (2002) establishes this point theoretically and also shows that enforcing best execution enhances competition between market places. Key to the economic intuition of this theoretical result is the following reasoning: With multiple trading venues, searching for the best price is costly. Brokers might be tempted to avoid incurring this cost. This would result in inferior execution for investors, and correspondingly large transactions costs. To restore brokers' incentives to engage costly search, financial penalties must be imposed by the regulator when detecting lack of compliance with best execution rules. Furthermore, if brokers do not search for the best price, competitors of the primary exchange (ATSs or intermediaries internalizing order flow) face less incentives to post competitive quotes to attract orders.

While the enforcement of best execution rules is desirable, it is possible only under certain conditions:

- It must be possible to measure execution quality precisely, and define a clearcut frontier between good conducts and bad ones. Like in the US, this measure could be based on the following two criteria: i) comparison with the prevailing price, and ii) promptness of execution. Some other factors such as the quality of information provided to the client or the quality of the counterparty may be taken into account on clients' request.
- Pre-trade transparency rules must be enforced. Indeed, regulators have to take into account, as the SEC does, the "circumstances" of the trade, including information on current quotes and depth. As pre-trade transparency

and access to all market venues (including firms internalizing their order flow) are enhanced, best execution gets easier to enforce and competition is spurred.

III.2.4) Some exemptions for block trading.

In the case of large trades and blocks, it might be appropriate to relax the regulations suggested above.

First, regulation is especially important, to protect investors rights, in the case of retail or medium sized investors, who may not have the time, skills or resources necessary to monitor financial intermediaries. In contrast, large traders are in a better position to monitor execution quality and bargain with intermediaries.

Second, executing large trades, such as blocks, is particularly delicate. Minimizing market impact for these trades may require some deviation from full pre-trade transparency and explicit non-discretionary order handling algorithms.

The market impact of large orders can be decomposed in two components:

- First, consider the case where a large buy order is immediately executed against the limit order book. As it walks up the book, the order is executed at higher and higher marginal prices. This impact is particularly severe if the order book is not very deep.
- Second, suppose the block is traded dynamically, relying on a sequence of relatively smaller market and limit orders. As the market observes this sequence of orders, it updates upward the probability that a professional investor, or the in-house research service of institutional investors, have identified positive information about the value of the share. In line with this updating, the ask quotes in the book are raised. (This is in line with the theoretical analysis of Kyle, 1985). This reduces the execution quality of the block trade. While this reasoning could in principle also be applied to small trades, it is more likely to be relevant for very large orders (as documented empirically by Seppi, 1992).

Mitigating this market impact calls for sophisticated strategies, involving, for example, order splitting, placing hidden orders, or possibly some telephone negotiations. Diligent and competent brokers and traders can be entrusted with the task of fine tuning such strategies.

It is reasonable to argue that market regulation should be adapted to make this possible. Correspondingly, we recommend that:

- An exemption to the pre-trade transparency requirement be admitted for very large trades. We propose a threshold of 500.000 euros: the experience of Euronext shows that below this cutoff, the central order book is suitable for almost all transactions: In a typical month (May 2002), only 0,4 % of trades inferior to 500.000 Euros were put through the market as crosses (“applications”).
- Fund managers should be able to demonstrate to investors that they minimize the market impact. Brokers should be able to demonstrate the same to asset managers. European regulators should require asset managers to systematically evaluate the execution performance of their brokers. The existence of electronic and centralized markets in Europe should facilitate precise comparisons between actual trades and market conditions at the time of those trades. Of course, such systems should be run by independent entities, not brokers.

References

- Admati, A., and P. Pfleiderer, 1988, "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies*, 1, 3-40.
- Auguy, Mathias, 2002, Internalization: Impact(s) on the process of price discovery, Working paper, MAConsultant, July 2002
- Barclay, M., and J. Wagner, 1993, "Stealth trading and volatility: Which trades move prices?" *Journal of Financial Economics*, 281-306.
- Barclay, M., W. Christie, J. Harris, E. Kandel, and P. Schultz, 1999, "The Effects of Market Reform on the Trading Costs and Depths of Nasdaq Stocks," *Journal of Finance*, 54, 1-34.
- Battalio, R., 1997, "Third market broker dealers: Cost competitors or cream skimmers? An empirical analysis," *Journal of Finance*, 341-352.
- Battalio, R., J. Greene, and R. Jennings, 1997, "Do Competing Specialists and Preferencing Dealers Affect Market Quality?" *Review of Financial Studies*, 10, 969-993.
- Battalio, R., and C. Holden, 2001, A simple model of payment for order flow, internalization and total trading costs, *Journal of Financial Markets*, 33-71.
- Bessembinder, H., and H. Kaufman, 1997, "A Cross-Exchange Comparison of Execution Costs and Information Flow for NYSE-Listed Stocks," *Journal of Financial Economics*, 46, 293-319.
- Biais, B., L. Glosten, and C. Spatt, 2002, "The microstructure of stock markets", Working Paper, Columbia University.

Biais, B., C. Bisiere and C. Spatt, 2002, Imperfect Competition in Financial Markets: Island versus Nasdaq, Working paper, Toulouse University.

Biais, B., P. Hillion, and C. Spatt, 1995, "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse," *Journal of Finance*, 50, 1655-1689.

Biais, B., P. Hillion, and C. Spatt, 1999, "Price Discovery and Learning during the Preopening Period in the Paris Bourse," *Journal of Political Economy*, 107, 1218-1248.

Biais, B., P. Bossaerts, and J. Rochet, 2001, "An Optimal IPO Mechanism," forthcoming, *Review of Economic Studies*.

Chen, H.S., and J. Ritter, 1999, The seven percent solution, *Journal of Finance*.

Christie, W., J. Harris, and P. Schultz, 1994, "Why Did NASDAQ Market Makers Stop Avoiding Odd-Eighth Quotes?", *Journal of Finance*, 49, 1841-1860.

Christie, W., and P. Schultz, 1994, "Why do NASDAQ Market Makers Avoid Odd-Eighth Quotes?", *Journal of Finance*, 49, 1813-1840.

Corwin, S., and M. Lipson, 2000, "Order Flow and Liquidity around NYSE Trading Halts," *Journal of Finance*, 55, 1771-1805.

Foucault, T., 2002, Best Execution and Competition between Trading Venues, EAMA, a collection of essays on Best Execution, 2002.

Glosten, L., 1994, "Is the Electronic Open Limit Order Book Inevitable?", *Journal of Finance*, 49, 1127-1161.

Goldreich, D., 2002, Uniform price Treasury auctions with heterogeneous information, Working Paper, London Business School.

Goldstein, M., and A. Kavajecz, 2000, "Eighths, Sixteenths and Market Depth: Changes in Tick Size and Liquidity Provision on the NYSE," *Journal of Financial Economics*, 56, 125-149.

Griffiths, M., B. Smith, S. Turnbull, and R. White, 2000, "The Costs and Determinants of Order Aggressiveness," *Journal of Financial Economics*, 56, 65-88.

Hansch, O., N. Naik, and S. Vishwanathan, 1999, "Preferencing, internalization, best execution and dealer profits", *The Journal of Finance*, 1799-1828.

Harris, L., and J. Hasbrouck, 1996, "Market vs. Limit Orders: The SuperDOT Evidence on Order Submission Strategy," *Journal of Financial and Quantitative Analysis*, 31, 213-231.

Hasbrouck, J., and G. Sofianos, 1993, "The Trades of Market Makers: An Empirical Analysis of NYSE Specialists," *Journal of Finance*, 48, 1565-1593.

Huang, R., 2002, "The Quality of ECN and Nasdaq Market Maker Quotes," *Journal of Finance*, 57.

Kavajecz, K., 1999, "A specialist's quoted depth and the limit order book," *Journal of Finance*, 747--772.

Kyle, A., 1985, "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315-1335.

Lee, C., B. Mucklow, and M. J. Ready, 1993, "Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis," *Review of Financial Studies*, 6, 345-374.

Lehman, B., and D. Modest, 1995, "Market Structure and Liquidity on the Tokyo Stock Exchange," in Andrew W. Lo (ed.), *The Industrial Organization and Regulation of Securities Markets* (Chicago: University of Chicago Press, 1995).

Naik, N., and P. Yadav, 1999, "The Effects of Market Reform on Trading Costs of Public Investors: Evidence from the London Stock Exchange," working paper.

Pagano, M., 1989, "Trading Volume and Asset Liquidity," *Quarterly Journal of Economics*, 104, 255-276.

Petersen, M., and D. Fialkowski, 1994, "Posted Versus Effective Spreads," *Journal of Financial Economics*, 35, 269-292.

Ready, M., 1999, "The Specialist's Discretion: Stopped Orders and Price Improvement," *Review of Financial Studies*, 12, 1075-1112.

Rock, K., 1990, "The Specialist's Order Book and Price Anomalies," working paper, Harvard University.

Seppi, D., 1992, "Block Trading and Information Revelation Around Quarterly Earnings Announcements," *Review of Financial Studies*, 5, 281-305.

Stoll, H., and R. Whaley, 1990, "Stock Market Structure and Volatility," *Review of Financial Studies*, 3, 37-71.

Walras, L., *Eléments d'Economie Politique Pure, ou Théorie de la Richesse Sociale*. Lausanne: Corbaz, 1874. 2d rev. ed. Lausanne: Rouge, 1889.