ELSEVIER
Journal of Economics and Business xxx (2005) xxx–xxx
Journal of Economics & Business

Slow and fast markets

Vassilis Polimenis *

A. Gary Anderson Graduate School of Management, University of California, Riverside, CA 92521-0203, USA

Received 7 September 2004; received in revised form 9 August 2005; accepted 2 September 2005

Abstract

Informed traders need liquidity in order to profit from their private information. Markets provide liquidity and are compensated by the information released through trading. Fast markets provide access to a limit order book. Slow markets provide execution in an auction-based trading floor. Hybrid markets combine both execution venues. It is shown here that the overall efficiency of a hybrid market is determined by its fast component. The introduction of a trading floor does not generate more informed trading, only takes trading away from the fast market. Trading floors are thus inherently competitive to the fast market. We provide conditions that determine the competitiveness of a trading floor with respect to a fast market.

Keywords: Slow market; Fast market; Traders

1. Introduction

Advances in communication and computing technologies are accelerating the tempo of introduction of successful new electronic markets. The traditional trading paradigm, that of an "open-outcry" auction-based trading floor, is changing to a more complicated and diverse order matching environment. Clearly, the introduction of new technologies brings promise for faster, less risky and more accurate trading executions. Electronic markets' representatives (and the SEC) believe that the current National Market System (NMS), much of which was designed and implemented in the 1970s and 1980s, needs a major overhaul to respond to the sweeping technological changes and the intensified inter-exchange competition.

Complete automation that eliminates latency in order execution, and creates perfect order timing sequencing, will probably make it difficult for exchange professionals to trade profitably

* Tel.: +1 951 827 3554; fax: +1 951 827 3970.
E-mail address: polimenis@ucr.edu.

0148-6195/$ – see front matter © 2005 Published by Elsevier Inc.
doi:10.1016/j.jeconbus.2005.09.002

in a decimalized world. As Peterffy and Battan[1] noted in the SEC's Market Structure Hearings held in New York on November 2002, "[d]esignated liquidity providers, therefore, have . . . to rely on their inherent time and place advantage in the manual market place – specifically, that they can see orders before others can see them and can take their time (sometimes up to 90 seconds) to decide whether to interact with those orders or not – in order to reap a reward for the services they provide".

Essentially, the inherent difference in the firmness of quotes posted in traditional "slow" floor-based exchanges gives them an advantage relative to their "fast" electronic competitors for liquidity. The key to understanding why is that a fast market quote is a true quote that an order can undoubtedly be filled against, while a slow market quote is more of an indication of a price. A slow market may be slow in updating its quotes, thus posting artificially attractive quotes, and blocking away electronic exchanges who cannot do so.

To address some of these problems the SEC in its recently published, newly re-proposed Regulation NMS differentiates markets (or quotes) into fast and slow. Essentially, electronic markets are thought of as fast markets, while manual, floor-based exchanges are considered slow markets. According to the proposal, fast markets will not be allowed to execute an order at a price that is inferior to an electronic market's best price, but will be able to trade through better posted prices on slow markets. That is, the newly proposed trade through rule will only protect quotations that are *immediately available* through automatic execution.

As a response, some exchanges are introducing (or enhancing existing) electronic-trading platforms so that they can still attract order flow by becoming essentially hybrid markets, thus retaining the advantages of the trade through rule. But these transformations have generated some debate within the floor-trading communities of brokers and specialists – the intermediaries who oversee the matching of buy and sell orders – who are at the heart of the auction-based price discovery process. One of the contested issues for the newly emerging hybrid exchanges is whether to allow for so-called "sweeping of the book". Sweeping refers to the ability to electronically execute transactions not only at the best price, but also at quotes above or below the best price.[2] It is argued that the electronic component creates in-house competition for the trading floor, since electronic access to the full limit order book would make it much more challenging for the floor community to co-exist within the newly emerging hybrid markets.

In the face of the significant developments outlined above, we need new theories that directly address issues related to competition between exchanges. Are hybrid markets good for price discovery? Do trading floors enhance the efficiency of the market? What changes in trading should we expect when a traditional exchange becomes hybrid by offering electronic access to its entire limit order book through sweeping?

Even though there is an extended literature on trading equilibria, it is not clear what the two canonical models of market microstructure described in Glosten and Milgrom (1985) and Kyle (1985) have to say about slow and fast markets. Kyle (1985) is among the first to model a batch-

[1] See also the recent Peterffy and Battan (2004) piece in *Financial Analysts Journal*.

[2] For example, NYSE recently proposed to upgrade Direct+® and allow sweeps. It should be emphasized that NYSE's sweep will be quite different from what we study here. Firstly, under the proposed plan, NYSE orders will sweep the book only up to the next "liquidity replenishment point". Liquidity replenishment points (LRPs) are predetermined price points at which the NYSE will briefly change to a non-automated market. But the most important difference is that in a NYSE sweep even though the order will be allowed to access the residual liquidity of the book, outside the published quote, it will only be price improved at the clean-up price. In this paper, we assume that sweeps are executed at progressive prices rather than the clean-up price.

auction market, where risk-neutral market makers compete and set a market clearing price by observing a noisy total order flow. In Glosten and Milgrom (1985) orders arrive and get executed individually, and the market makers protect themselves against the possibility of trading with an informed trader by properly managing their bid–ask spreads. Recently, Back and Baruch (2004) have gone a long way towards reconciling continuous time versions of the two models, by showing how equilibria of the Glosten–Milgrom-type converge to the Kyle equilibrium.

Despite their indisputable significance, the above equilibrium analyses do not lend themselves for a study of fast versus slow markets because there is no real element of delay. In these models, market efficiency guarantees that price immediately reflects all information available to the market makers. Slow markets are probably inefficient to some degree and we need a model that incorporates inefficiency if we are to capture "slowness".

In Kyle (1985) the sensitivity of an *efficient* market, λ, is endogenously specified as a signal-to-noise ratio. Polimenis (2005) provides the mechanism by which the exogenous sensitivity[3] for an *inefficient* market is corrected for execution delays by a risk averse insider who has to cope with the uncertainty of his order execution. The model of the slow market here is borrowed from Polimenis (2005), but here the trader's options are extended to allow for the choice to participate, up to an endogenously chosen degree, in a fast market.

On exchange competition, Glosten (1994) finds that the electronic limit order book dominates any other exchange; if the book does not offer liquidity for a trade, any other anonymous exchange would lose by staying open for trade. Seppi (1997) and its extension in Parlour and Seppi (2003) present a model of competition between a pure and a hybrid market. Parlour and Seppi find that it is unclear whether aggregate liquidity will increase or decrease relative to a single market depending on how the investor trades when he/she is indifferent. Santos and Scheinkman (2001) ask whether competition will eventually lead exchanges to "lower" their standards on who they allow to trade and how much. In their model, a monopolist makes trading more expensive and thus traders trade less and have a reduced incentive to default; this economizes collateral. Viswanathan and Wang (2002) also study the problem of hybrid versus pure markets and find that risk-averse traders may prefer a hybrid market.

Although the paper here is also related to exchange competition, it extends the literature by looking at competition from a new angle, namely the question of speed of execution, and the risk related to slow executions. In contrast, in the existing literature there is no speed dimension. For example, in the Seppi (1997) model, and its simple extension to a hybrid market in Parlour and Seppi (2003), the difference between a hybrid and a pure market is not due to execution speed, but due to the existence of a specialist who sets the price; trades still clear immediately at deterministic prices.

At a first blush, it would seem that, by fostering competition for liquidity among markets, having another trading venue would generate more informed trading. On the contrary, by using an idealized model of an open electronic book that allows unlimited sweeping (and assuming away a number of frictions and trading costs, which may be significant), we find that a slow market only "steals" away liquidity from a fast market, that is, the two markets compete for the same order flow. The degree of effectiveness of the trading floor in competing with the electronic book is determined by their depth differential. Limit orders placed in the book are safer for the insider but riskier from the liquidity providers' point of view; this explains why liquidity accessed through

[3] In Polimenis (2005) ρ denotes market sensitivity. To enhance readability, here we use Kyle's lambda notation for sensitivity.

sweeping the book can be more expensive (smaller depth) and still attractive. More specifically, it is shown that when competing for order flow on a stock with an electronic market λ_F^{-1} shares per dollar deep, a trading floor will be able to divert trades away from its electronic competitor only if it is at least twice as deep, $\lambda_o^{-1} > 2\lambda_F^{-1}$.

The first task of the paper is to present a formal model of the liquidity and price discovery mechanisms on both the trading floor as well as the hybrid markets that combine a floor with an electronic sweeping facility. Then, the paper focuses on the actions of a large informed trader who faces two uncertainties: how much information will they release and how much will this information affect their order execution? Liquid markets are characterized by large numbers of uninformed trades that effectively allow informed traders to trade without releasing much information. Equivalently, liquid markets can absorb orders fast, thus allowing informed traders to capture most of the, privately known, misvaluation of a security before their private information becomes public. For a short period, before her proprietary information becomes public, an informed trader is a liquidity monopsonist. In return for releasing information in the market, through her trades, the informed trader is compensated with liquidity supplied by uninformed traders.

Unlike the instantaneous and riskless execution offered by electronic markets, auction-based trading floors are characterized by “slow” executions. A slow order execution is undesirable for an informed trader, not due to the time value of money “tied” to the order, but rather, due to informational reasons. Specifically, a slow execution is risky because it exposes the block order to the market for larger periods of time, thus increasing informational costs. Mathematically, the price impact of the order is subordinated to the execution of the order, i.e. the order is impacting the market while it is being executed. Essentially, execution time becomes the medium for dissemination of information to the market. By prolonging executions, illiquid markets give more time to the market to “distill” and reflect private information, thus limiting the informed agent’s profits. Specifically, on the trading floor agents face:

- uncertainty in execution delays that grows with the order size, and declines with the liquidity of the floor, and
- uncertainty in price impact that grows with size, and declines with liquidity and depth.

Another critical issue, which differentiates trading on the floor from trading on an electronic platform, is the front-running and liquidity withholding behavior around a block trade. The paper presents some analysis of the different trading behavior in a traditional trading floor, one where the identity of the client, and thus her total interest, is transparent to the floor, and a trading floor that provides anonymity. As is shown in Polimenis (2005), in the anonymous floor, traders protect their trades from front-running, by spreading their orders into many smaller size sub-orders and actually trade more. This way, their early sub-orders get to be executed in an environment that has not yet reflected the full block order. Essentially, in an anonymous market, the liquidity providers who fill the early sub-orders have no way of knowing the true order size, and are price discriminated against. It is shown here that anonymous floors are indeed more attractive, and actually can attract trades by only being marginally deeper than the electronic book.

As in Polimenis (2005), the paper is not answering the question of who offers liquidity, why and at what price. By being a partial equilibrium,[4] the results are of limited scope, and refer to an

[4] All theoretical microstructure papers known to the author are partial equilibrium results. This is not due to lack of modelling skill, but rather a testament to the enormous complexity of the trading process.

idealized framework that assumes away many real-life frictions—speed limits in computing and communications technologies, limited bandwidth, errors, etc. Nevertheless, it takes a step in the right direction by modeling fast and slow markets, and providing simple trading rules that show how will large traders split their orders between the two types of markets. If we know how traders route their orders among fast and slow markets, we can also derive competitiveness criteria, for the two types of markets, based on parameters, such as the liquidity of the floor, and the depth of the limit order book, that can be easily calibrated by observing real time data available to large traders.

The critical, and probably strong, assumption of the model is that of subordination of price discovery to liquidity discovery on the trading floor; while the search for liquidity continues, prices continually get updated. The good news is that the specific type of subordination is of less importance; what is really needed for most derivations here is that, conditional on the liquidity of the floor, the price impact of an order is independent of the price impact of previous orders and only depends on this order's size. One may actually derive all the important intuition (minus the closed form end-results) with any type of subordination; the specific Normal Inverse Gaussian is only due to the assumption that both liquidity and price discovery processes follow diffusions, and provides tractable analytical results.

In Section 2, the liquidity and price discovery processes are explained for both slow and fast markets. In Section 3, optimal trading behavior for standalone trading floors is calculated. Finally, in Section 4, hybrid markets are discussed.

2. Price discovery in fast and slow markets

Fast markets are characterized by certain and immediate executions, i.e. when sweeping the totally transparent book, the agent knows her order will be immediately executed and at what price. As usual, private information has value because it indicates that securities are actually traded away from their fundamental value. According to her private information, the trader observes that, at time zero, the stock trades at a price that deviates from its intrinsic value by ΔP dollars. If the trader was a price taker, the existence of a non-zero ΔP would be an arbitrage opportunity. In reality, even though $\Delta P > 0$ will still lead to trading, there are limits to the profitability of the informed trader. Trading a small quantity of misvalued shares almost ensures a profit, but when order size grows, risky execution impact limits the benefit of private information.

A large informed trader knows that she is not a price taker. The components that determine the profitability of her strategy are the dollar value of the private information, ΔP, which captures the gain due to the current misvaluation of the security, and the price impact of the released information (cost to trade), I.

Limit order books are characterized by their breadth, λ_F,[5] measured in dollars of impact per transacted share. That is, λ_F is a sensitivity parameter; it captures the sensitivity of the fast market to trading. A trader who issues an order to a fast market with a sensitivity λ_F will *deterministically* impact the price by

$$I_F = \lambda_F q \tag{1}$$

The inverse, λ_F^{-1}, measured in shares per dollar, is a natural proxy for market depth, and captures the density of shares, placed by limit orders, in the order book.

[5] F stands for fast market.

2.1. Liquidity and price discovery on the floor

Slow markets, as their name implies, are characterized by uncertain and slow executions, i.e. when sending her order to a trading floor, the agent does not know when her order will be executed and at what price. The fundamental difference between the sensitivities of a fast and slow market is that the latter refers to a risky execution and thus needs to be corrected for risk. Specifically, on a fast market the sensitivity of the book, λ_F, directly provides the magnitude of price impact for a given order (1). On a floor, the sensitivity, λ_o, provides *only an expectation*; a trader who submits an order for q shares to the floor expects an impact equal to $\lambda_o q$. Polimenis (2005) introduces a method that corrects a λ for risk, and we briefly review the model here.

The central doctrine in Polimenis (2005) is that a slow order execution is undesirable for an informed trader, not due to lost interest in the money "tied" to the order, but rather, due to informational reasons. In the inefficient markets modeled there, information is not instantaneously reflected in prices, and thus slow executions are risky because they expose orders to the market for a prolonged period of time and thus increase the informational costs. Essentially, execution time becomes the medium for dissemination of information to the market. Illiquid floors prolong executions and allow more time for the market makers to absorb and reflect private information, thus limiting the informed agent's profits.

Of course, if we are to model illiquid markets in a realistic way, execution delays will have to depend on both the liquidity of the market as well as the size of the particular order. In liquid trading floors, the large number of uninformed traders creates noise and allows an informed trader to release less information per share traded. Liquid markets are characterized by the large number of shares that can be traded before private information becomes public. Large orders will be expected to take more time, thus revealing more information, than small ones. Since time connects order sizes to information, the liquidity, l_o, of the floor is defined as the expected rate at which orders are absorbed, and is measured in transacted shares per second.

The liquidity discovery process is a Brownian motion B_t with drift $l_o > 0$, and the order is completed after τ seconds, where τ is the stopping time defined as

$$\tau = \min\{t : B_t = q > 0\}$$

In this case, the trader expects that her order will take $E\tau = \frac{q}{l_o}$ (seconds) to get executed on the floor. While execution takes place, new information about the security is revealed. Such information gets reflected in the execution price, and limits the profitability of the informed trade from ΔP to $\Delta P - I(q)$, where $I(q)$ is the total dollar impact of information released during the time, τ, it takes for the order to get executed.

We now come to the critical assumption about price discovery on the floor, i.e. the mechanism by which information gets revealed and priced. Mathematically, price discovery is subordinated to liquidity discovery, i.e. prices are impacted for as long as the order gets executed. Formally, the impact from a q-size order is

$$I(q) = I(\tau_q) \tag{2}$$

The final observation is that since the floor's liquidity is measured in shares per second and since the floor's lambda, λ_o, is measured in dollars per share, their product

$$v_o = \lambda_o \times l_o \tag{3}$$

measured in dollars per second, captures the expected drift of the price impact. That is, the price drifts "against" the trader at a rate υ_o. To capture the stochastic property of the market, impact is modeled as

$$I(q) = \upsilon_o \tau_q + W(\tau_q) \tag{4}$$

where τ_q is the stopping time defined above and $W(t)$ is the usual Brownian motion.

If during the order execution in a slow market, new information is expected to move the market by υ_o dollars every second, the expected total impact of the trade will be[6]

$$\mathrm{EI_S} = \upsilon_o \times \left(\frac{q}{l_o}\right) = \lambda_o \times q \tag{5}$$

Observe that the efficient Kyle-type floor (fast market) is taken as the limit of an infinitely liquid slow floor, $l=\infty$. In this case, the impact drift (3) would also be infinite, i.e. prices would adjust instantaneously and the impact (4) would become deterministic, $I_S = \mathrm{EI_S}$. It is thus clear that, in our model here, the defining characteristic of a slow market is its finite liquidity. Notice that, keeping λ_o fixed, liquidity does not affect expected price impact (5). Thus, a risk neutral trader (as in Kyle) does not care about slow versus fast markets. Nevertheless, a liquid security will be preferred by a risk averse trader because, by accelerating execution, liquidity lowers price impact risk

$$\mathrm{Var}(I_S) = \frac{q}{l_o}(1 + \lambda_o^2) \tag{6}$$

The total impact of the submitted order follows a Normal Inverse Gaussian law[7]

$$I(q) \sim \mathrm{NIG}(\alpha, \upsilon_o, 0, q) \tag{7}$$

with $\alpha = (\upsilon_o^2 + l_o^2)^{1/2}$.

It is important to recognize that the impact follows a Lévy process with respect to trading q, the only quantity the trader controls. That is, the impacts of successive trades are independently distributed.

3. Trading on the floor

As discussed previously, to compare fast versus slow markets we need an agent that cares about execution delays and the risk associated with them. The trader here deals with risk by exhibiting a constant absolute risk aversion η, and is able to observe the markets, i.e. knows the lambda and liquidity offered by the floor, λ_o and l_o, respectively, so that she can calculate the impact drift for her orders, υ_o.

In a fully transparent trading floor, everybody knows who he/she is trading with and the informed trader cannot conceal her identity.[8] Thus, in such markets there is no point in breaking the order into smaller sub-orders and submitting them separately. Floor specialists will know the

[6] S for slow market.

[7] For a more complete discussion of the Normal Inverse Gaussian, see Barndorff-Nielsen (1998).

[8] With respect to trader identity, the two cases studied here are rather extreme (complete anonymity or full transparency). Real cases will be somewhere between.

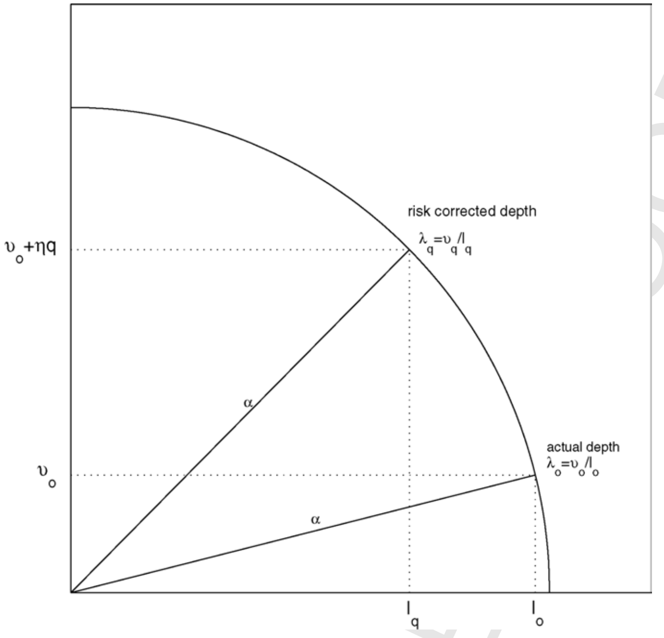

Fig. 1. Polimenis (2005) provides the method to correct for risk the sensitivity of an inefficient market (slow market here).

issuer of the sub-orders and treat them as a single big order, by front-running it and withholding liquidity.

When a block order is submitted to the trading floor, the block will be transacted at a single price that fully reflects the entire impact of the information released during the order execution. The utility – a trader with initial wealth W gets – from issuing a single block order of size q equals

$$U_{\mathrm{S}} = E^{q} - \mathrm{e}^{-\eta W}\,\mathrm{e}^{-\eta q(\Delta P - I_{\mathrm{S}}(q))} \tag{8}$$

where the q superscript in the expectation operator explicitly shows that the expectation is taken conditionally on the order size, since q determines size, as well as per share profits. The trader will buy underpriced and sell overpriced securities, and without loss of generality, we assume that ΔP and q are positive.

Taking iterated expectations, by conditioning on the information τ, and using the moment generating function for the inverse Gaussian law

$$E^{q}\,\mathrm{e}^{s\tau} = \mathrm{e}^{q(l_{\mathrm{o}} - \sqrt{l_{\mathrm{o}}^{2} - 2s})} \tag{9}$$

we find that the utility gain from block trading in a slow market equals

$$G^{\mathrm{TS}} = \ln \frac{U_{\mathrm{o}}}{U_{\mathrm{S}}} = \eta q\,\Delta P - q(l_{\mathrm{o}} - l_{q}) \tag{10}$$

where $l_q = \sqrt{\alpha^2 - \upsilon_q^2}$ with $\upsilon_q = \upsilon_{\mathrm{o}} + \eta q$ (Fig. 1).

Unlike Kyle's model (fast market), where the insider derives monopolistic power from setting the price through q, the insider here, when trading in a slow market, also sets the price of the floor liquidity (her cost to trade) as[9]

$$P_q^{\mathrm{TS}} = l_{\mathrm{o}} - l_q \tag{11}$$

For a short time period,[10] being the sole owner of her proprietary information, the informed trader acts as a liquidity monopsonist who buys liquidity (uninformed shares) and sells information (informed shares). As any monopsonist, she is confronted with an increasing supply curve for liquidity

$$\frac{\mathrm{d}P}{\mathrm{d}q} = \eta\lambda_q > 0 \tag{12}$$

where $\lambda_q = \frac{v_q}{l_q}$, while the marginal per share benefit is fixed at $\eta\,\Delta P$. The Polimenis (2005) model for liquidity correction properly predicts that if the trader is small, the price of liquidity, P_0, is zero; a small investor is a price taker since she does not have to pay for liquidity

$$l_0 = l_{\mathrm{o}} \tag{13}$$

3.1. The optimal trade

In transparent markets, the informed agent's problem is to decide for the optimal block order size, q_{TS}. Given (10), a trader who trades a single block in a trading floor, solves

$$q_{\mathrm{TS}} = \arg\max_q \eta q\,\Delta P - qP_q \tag{14}$$

As a liquidity monopsony, the informed trader trades to the point that equates her constant marginal revenue

$$\mathrm{MR} = \eta\,\Delta P = \mathrm{AR} \tag{15}$$

to her increasing marginal liquidity cost

$$\mathrm{MC} = P_q + q\frac{\mathrm{d}P_q}{\mathrm{d}q} > \mathrm{AC} = P_q \tag{16}$$

The first-order condition for the optimal q, MR = MC, translates to (see Fig. 3)

$$P_{q_{\mathrm{TS}}}^{\mathrm{TS}} + \eta\lambda_{q_{\mathrm{TS}}} = \eta\,\Delta P \tag{17}$$

3.1.1. Price differentials

When the stand-alone book and the floor post the same prices, and are characterized by the same λ, since trading on a floor is risky, the trader will clearly prefer to trade on the book. As we will see later, in a hybrid market that offers both an electronic sweeping facility and a floor with common pricing, the trader will choose to divert some trading on the risky floor if the floor presents a better λ. Before we deal with this question, an easier question is when *a stand-alone floor* will improve the agent by simply offering a better price.

[9] TS for transparent slow market.
[10] Before her private information becomes public.

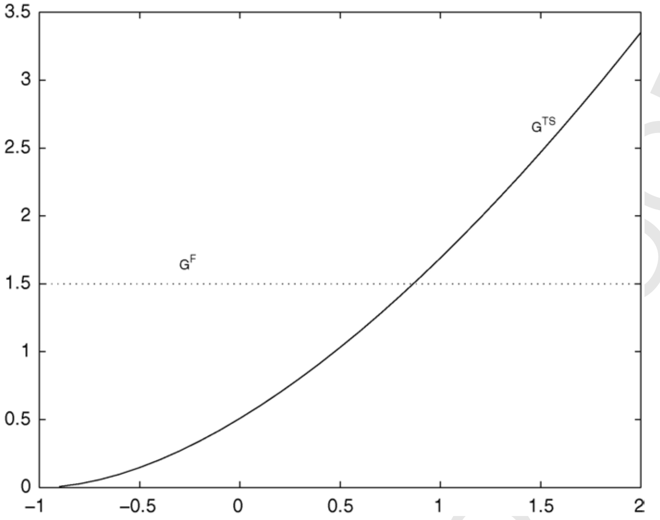

Fig. 2. Trading on the floor can only improve utility if the floor posts a better price than the electronic book. Here, we plot the total utility G^{TS} from selling on the stand-alone floor as a function of the price differential between the floor and the book. With $\lambda_{\mathrm{F}} = \lambda_{\mathrm{o}} = (¢1)/1000$ shares, floor liquidity $l_{\mathrm{o}} = 10^6$ shares per day and $\Delta P = ¢1$, an agent with $\eta = 3$ will choose the floor only if it offers a better price by ¢0.88 over the book.

Since we want to focus on price differentials between the two facilities while keeping everything else constant, let us assume for now only that $\lambda_{\mathrm{F}} = \lambda_{\mathrm{o}} = \lambda$. Using (17), we can recover the total utility gain extracted by the insider who trades solely on the floor

$$G^{\mathrm{TS}} = \eta \lambda_{q_{\mathrm{TS}}} q_{\mathrm{TS}}^2 \tag{18}$$

On the other hand, on the book the agent will clearly trade $q_{\mathrm{F}} = \Delta P/\lambda$, and improve his utility by $G^{\mathrm{F}} = \eta \lambda q_{\mathrm{F}}^2/2$. The two formulas are not directly comparable, because the trader, faced with the price uncertainty on the floor, will trade more on the book, $q_{\mathrm{F}} > q_{\mathrm{TS}}$.

Nevertheless, there is a price differential, that depends on the floor liquidity, l_{o}, and the agent's η, that will induce the trader to prefer the floor (Fig. 2).

In the example of Fig. 3, an ¢0.88 differential will improve the agent. Thus, if the floor posted a better price by less than ¢0.88 and the trade through rule forced the agent to trade there, the agent's position would be worsened.

3.2. Trading efficient markets

By definition, the markets here are inefficient from an informational point of view in that they do not reflect all the available public information as in Kyle or Glosten and Milgrom. Despite the undisputed significance, informational efficiency in their results heavily relies upon a very specific structure of the inside information; the inside information is either a zero-one value or a zero-mean normal with known variance. It is only such strict assumptions about the missing inside information that allow market makers to tractably form the efficient price as an expectation given their publicly available information. In realistic situations, the unwillingness of market makers to impose such a heavy structure on what they do not know about the fundamental value may lead

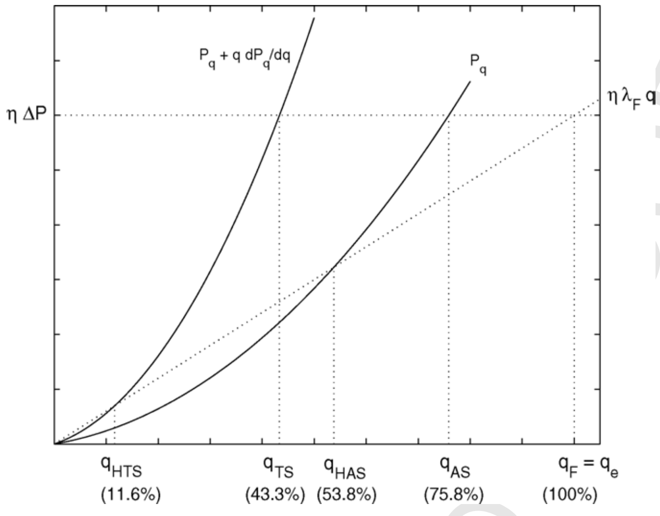

Fig. 3. Trading on the trading floor as a percent of the total efficient trading, q_{e}. In a stand-alone fast market $q_{\mathrm{F}} = q_{\mathrm{H}}$. Anonymity on the trading floor leads to more trading ($q_{\mathrm{AS}} > q_{\mathrm{TS}}$ and $q_{\mathrm{HAS}} > q_{\mathrm{HTS}}$). In a hybrid market, trading on the floor is given by the intersection of the $\eta\lambda_{\mathrm{F}}q$ line with the proper marginal floor liquidity cost (P_q for anonymous floors and $P_q + q\frac{\mathrm{d}P}{\mathrm{d}q}$ for transparent trading).

them to correct prices in a more mechanical way. Nevertheless, by correcting prices in the right direction, markets still tend to become more efficient by attracting more informed trade.

It is easy to see that as a liquidity monopsonist the informed trader buys less liquidity thus revealing less information in the markets. But how much informed trading, q_{e}, is needed to bring the market to its full informational efficiency? To understand this we need an operational definition of a trading efficient market. In such markets informed trading has reached its maximum profitable amount; it is the best we can hope for, given that informed traders are rational profit making agents. Equivalently, a market has reached its trading efficient state when there are no more utility gains from re-distributing the private information. Specifically,

Definition 1. A market is trading efficient when a new trader, endowed with the insider's information, cannot profitably trade.

Let us start with a market where informed traders have already submitted trades for q shares. Since a new trade will have to be executed in a floor that already works on executing market orders for q shares, the post-trade utility for a new trader j who has been endowed with the information equals

$$U(q_j) = U_{\mathrm{o}} E\, \mathrm{e}^{-\eta q_j(\Delta P - I(q+q_j))} \tag{19}$$

The critical observation is that, as a Brownian motion subordinated to the liquidity discovery Brownian motion, price discovery (4) is a Lévy process with respect to trading q. A Lévy process is characterized by increments that are independent and identically distributed. Thus, given the trade q, the utility gain equals

$$G = \ln \frac{U_{\mathrm{o}}}{U(q_j)} = \eta q_j\, \Delta P - (q+q_j)P_{q_j} \tag{20}$$

Eq. (20) points to a fundamental and counter-intuitive characteristic of the market for liquidity as it is modeled here. In usual markets, the price of a commodity is determined by the total demand, and the total cost for the price taking agent equals the product of her demand times the market-wide clearing price, $q_j \times P_{q+q_j}$. In the market for liquidity, each agent pays an individual price totally determined only by their trade P_{q_j}, but their total cost depends on the entire liquidity demand, $(q+q_j) \times P_{q_j}$. Essentially here, it is as if liquidity markets differentiate agents by the amount they trade. Unlike "small" traders, agents who trade large quantities are informed and they pay a large price. Anonymous markets, that we will study next, to the benefit of informed traders cannot differentiate them by their size.

From (20) the marginal gain of the new trader is

$$\frac{\mathrm{d}G}{\mathrm{d}q_j} = \eta\,\Delta P - P_{q_j} - (q+q_j)\left[\frac{\mathrm{d}P}{\mathrm{d}q}\right]_{q_j} \tag{21}$$

Since, at full informational efficiency, no trader can benefit if given the inside information, the amount of insider trading, q_{e}, that will bring the market to its trading efficient state is such that the initial marginal gain for a new trader has to be zero

$$\left[\frac{\mathrm{d}G}{\mathrm{d}q_j}\right]_0 = \eta\,\Delta P - q_{\mathrm{e}}\left[\frac{\mathrm{d}P}{\mathrm{d}q}\right]_0 = 0 \tag{22}$$

where we used the fact that $P_0 = 0$. From (12) we have $[\frac{\mathrm{d}P}{\mathrm{d}q}]_0 = \eta\lambda_{\mathrm{o}}$, and we recover

$$q_{\mathrm{e}} = \frac{\Delta P}{\lambda_{\mathrm{o}}} \tag{23}$$

Given the dynamics of the trading and price transmission mechanisms, the above condition is no surprise; the amount of maximum insider trading is inversely related to the total sensitivity of the market, and directly proportional to the value of the information that needs to be disclosed. Deep trading floors generate more inside trading.

3.3. *Trading in anonymous floors*

When a large block order arrives on the trading floor, specialists and floor brokers will try to front-run the block, in an effort to benefit from its impact, or if they want to take the opposite side, will avoid trading, thus refusing critical liquidity to the block. Essentially, by doing so, other traders will refuse to be price discriminated against by providing liquidity at the early stages of a block transaction. It is to the benefit of informed traders, when trading in an anonymous floor, to limit their exposure by not showing the full extend of their interest in their orders. This is accomplished by spreading their block trades into smaller trades.

As is shown in Polimenis (2005), when breaking the block order into N sub-orders, the trader benefits from only exposing small orders of size q/N to the floor. Even though all orders are submitted at the same time, and executed at a random order, sub-orders that are executed early are executed in a floor that has not yet reflected the informational impact of the entire block. As the sub-orders get consecutively executed, the later orders will be executed in an environment that becomes progressively more impacted. If the sensitivity and liquidity of the anonymous trading floor are not changing, the average liquidity cost is the arithmetic average of the increasing liquidity supply curve on the floor

$$\mathrm{AC} = \frac{1}{N}\sum_{i=1}^{N} P_{i_q/N} \tag{24}$$

and thus, clearly smaller than the average cost of a block, P_q. The agent gains from submitting N small anonymous orders instead of submitting a single block order.

The agent pays the smallest liquidity premium by breaking to and submitting infinitesimally small orders. If the optimal total trade size for such a series of infinitesimal orders is q_{AS},[11] the average gain equals

$$G^{\mathrm{AS}} = \eta q_{\mathrm{AS}}\,\Delta P - \int_0^{q_{\mathrm{AS}}} P_q\,\mathrm{d}q \tag{25}$$

Essentially, in anonymous floors, the agent can price discriminate against early liquidity providers who have no way of knowing who submitted the order.

3.4. Anonymous markets are more efficient than transparent markets

It is thus clear that when given the ability to minimize the liquidity cost, by spreading her orders the trader solves

$$q_{\mathrm{AS}} = \arg\max_q \eta q\,\Delta P - C^{\mathrm{A}}(q) \tag{26}$$

where the cost of anonymous liquidity is

$$C^{\mathrm{A}}(q) = \int_0^{q} P_y\,\mathrm{d}y \tag{27}$$

The first-order condition for the optimal total trade quantity which has been spread into infinitesimal sub-orders, q_{AS}, is

$$\eta\,\Delta P = P_{q_{\mathrm{AS}}} \tag{28}$$

That is, the optimal trade in any anonymous market[12] equates the marginal benefit of an extra share to the price for liquidity in that market. When, by hiding one's identity, the true trading interest is not revealed, the marginal cost of an extra share traded is only the cost of trading this share, since the liquidity providers who traded with the trader so far have no way of knowing that this extra share will be traded. This way, early liquidity providers are price discriminated against.

By comparing with the first-order conditions for the transparent floor, q_{TS}, the optimal single block order in (17), we see that

$$P_{q_{\mathrm{AS}}} = P_{q_{\mathrm{TS}}} + q_{\mathrm{TS}}\left[\frac{\mathrm{d}P}{\mathrm{d}q}\right]_{q_{\mathrm{TS}}} > P_{q_{\mathrm{TS}}} \tag{29}$$

and since, P_q is a monotonically increasing function of q, we see that when offered anonymity a trader will always trade more.

[11] AS for anonymous slow market.
[12] Both fast and slow.

4. Sweeping the book

Since by definition, the book displays limit orders, a model that completely addresses the question of trading on the floor versus sweeping the book would have to endogenously resolve the question of when traders decide to submit limit orders instead of market orders, and if so at which price point. Even though Polimenis (2005) offers some criterion as to when traders would decide to submit limit orders, the discussion is not complete enough for our purposes here; here, we need the density of the book at each price point.

In order to reach analytically tractable solutions, we assume a constant depth book, a strong assumption since books tend to get thinner away from the current price point. Thus, here, the limit order book is a fast market, and, sweeping the book is equivalent to accessing a fast market where shares are executed immediately and with no risk. The important quantity of a limit order book is the depth λ_F^{-1}. When sweeping q shares from the book, the agent releases information and immediately impacts the price by $\lambda_F q$. Since sweeping the book is an inherently anonymous operation,[13] that is by its nature the electronic limit order book is "discriminatory", when the insider submits an order through the book, her gain is

$$G^F = \ln \frac{U_o}{U} = \eta q \,\Delta P - \int_0^q \eta \lambda_F y \,\mathrm{d}y \tag{30}$$

Thus, the price for liquidity in the fast market is

$$P_q^F = \eta \lambda_F q \tag{31}$$

and the trader will trade q_F shares, where

$$\Delta P = \lambda_F q_F \tag{32}$$

pushing the *stand-alone* fast market to efficiency since at the end of trading the entire ΔP will be reflected in the price. Observe that even though the trader is risk averse, when sweeping the book there is no risk involved and η drops out.

4.1. Trading on the floor after sweeping the book

In a hybrid market, where a trader has access to a trading floor and the entire limit order book, she has one more control variable. Namely, she can choose how many shares to sweep from the book, q_{HF}, before she sends her remaining order, q_{HS}, to the floor.[14] If $q_H = q_{HF} + q_{HS}$ is the entire trade size,

$$i = \frac{q_{HF}}{q_H} \tag{33}$$

denotes the fraction of her order the trader chooses to sweep the book for, and $(1-i)$% will be routed through the floor. The first i% shares will be executed instantaneously, while the remainder $(1-i)$% will be executed after a delay $\tau_{q_{HS}}$. The parameter i thus captures the immediacy of the order.

[13] This point is made very clear in Glosten (1994).

[14] HF and HS for trading on the fast and slow components of a hybrid market, respectively.

As we saw previously, when a block order is submitted to a transparent floor, the block will be transacted at a single price that fully reflects the entire impact of the information released during the order execution. When allowed to sweep the book, the trader has a choice. She may choose how deeply (i.e. how far from the current market) to sweep the book. When she sweeps deeply, she will have a larger portion of her order being executed immediately. On the other hand, choosing not to sweep that deeply, but rather to wait through her order being executed against incoming uninformed trades, she is incurring large delays and possibly large impact through exposing her order for longer periods.

The total impact of a hybrid trade is decomposed into two different sources: a direct component because of sweeping and an indirect due to private information leaking and becoming public during execution on the floor.

The total utility gain, G^{HT},[15] from trading in a fast electronic market, and a *slow but transparent* trading floor, equals

$$G^{\mathrm{HT}} = \eta q_{\mathrm{H}}\, \Delta P - \int_0^{q_{\mathrm{F}}} \eta\lambda_{\mathrm{F}} q\, \mathrm{d}q - \eta\lambda_{\mathrm{F}} q_{\mathrm{F}} q_{\mathrm{S}} - q_{\mathrm{S}}(l_{\mathrm{o}} - l_{q_{\mathrm{S}}}) \tag{34}$$

Essentially, the trader buys the liquidity to trade in two different places. Sweeping the book happens instantaneously, and the liquidity cost of the fast component in a hybrid market is the same as the liquidity cost should the fast market operate alone (not as part of a hybrid market)

$$P_q^{\mathrm{HF}} = P_q^{\mathrm{F}} = \eta\lambda_{\mathrm{F}} q \tag{35}$$

On the other hand, since sweeping the book releases information immediately, this information will also be immediately reflected in trading ensuing on the floor. Thus, the liquidity price in the slow market (trading floor) will also depend on the amount traded on the fast market (limit order book)

$$P_{q_{\mathrm{S}}}^{\mathrm{HS}} = P_{q_{\mathrm{F}}}^{\mathrm{F}} + P_{q_{\mathrm{S}}}^{\mathrm{S}} \tag{36}$$

Taking the derivative of (34) with respect to the amount of fast trading, q_{F}, we find that the amounts traded in the fast and slow markets have to combine as follows:

$$\Delta P = \lambda_{\mathrm{F}} q_{\mathrm{H}} \tag{37}$$

Combining (32) and (37), we see that:

Lemma 1. *Hybrid markets do not generate more informed trading than stand-alone fast markets*

$$q_{\mathrm{H}} = q_{\mathrm{F}} \tag{38}$$

This lemma shows that the existence of the trading floor does not enhance the overall informational efficiency. When it co-exists with a fast market, a slow market will only “steal” liquidity without offering any informational benefits. Lemma 1 is a negative statement for the existence of the so-called hybrid markets that combine electronic and auction-based execution facilities. Finally, Lemma 1 predicts something many market participants have asserted all along: the two types of markets are inherently competitive since they will have to split the same liquidity.

Corollary 1. *In a hybrid market informational efficiency is determined by the depth of the fast market.*

[15] HT for hybrid with transparent floor.

This corollary implies that the sensitivity of the slow markets will not play any role in total trading, and the overall efficiency of the hybrid. This is totally determined by the fast market. As the next lemma shows though, the sensitivity of the slow market will determine the amount of trading that will be diverted to the floor.

Having solved the problem of the total trading in a hybrid market, $q_{\mathrm{H}} = \frac{\Delta P}{\lambda_{\mathrm{F}}}$, we move to the next important question: will there be any trading on the floor, and if so how much? Rewrite (34) as a function only of the trading on the floor, q_{S},

$$G^{\mathrm{H}}(q_{\mathrm{S}}) = \eta q_{\mathrm{H}}\,\Delta P - \int_0^{q_{\mathrm{H}}-q_{\mathrm{S}}} \eta\lambda_{\mathrm{F}} q\,\mathrm{d}q - \eta\lambda_{\mathrm{F}}(q_{\mathrm{H}} - q_{\mathrm{S}})q_{\mathrm{S}} - q_{\mathrm{S}}(l_{\mathrm{o}} - l_{q_{\mathrm{S}}}) \tag{39}$$

and take the derivative of G^{H} with respect to q_{S}

$$\frac{\partial G^{\mathrm{H}}}{\partial q_{\mathrm{S}}} = \eta\lambda_{\mathrm{F}} q_{\mathrm{S}} - (l_{\mathrm{o}} - l_{q_{\mathrm{S}}}) - \eta\lambda_{q_{\mathrm{S}}} q_{\mathrm{S}} \tag{40}$$

Since

$$\left[\frac{\partial G^{\mathrm{H}}}{\partial q_{\mathrm{S}}}\right]_{q_{\mathrm{S}}=0} = 0 \tag{41}$$

the trader will only trade at the floor when the second derivative at zero is positive. Doing the algebra, condition

$$\left[\frac{\partial^2 G^{\mathrm{H}}}{\partial q_{\mathrm{S}}^2}\right]_{q_{\mathrm{S}}=0} > 0 \tag{42}$$

leads to the following competitive criterion:

Lemma 2. *In a hybrid market with a transparent floor, trading will be diverted from the fast to the slow market only if*

$$\lambda_{\mathrm{o}} < \frac{\lambda_{\mathrm{F}}}{2} \tag{43}$$

The optimal trade condition for q_{HTS} is depicted in Fig. 3.

4.2. Hybrid markets with anonymous trading floors

Anonymous floors are more competitive than transparent. It is thus easier for an anonymous floor to take liquidity from the fast market. If the trading floor of the hybrid market allows for anonymous trading, the informed trader will break her block into many small trades. In this case, as we saw previously, early liquidity providers on the floor will be price discriminated, and the total utility gain[16] for the informed trader equals

$$G^{\mathrm{HA}} = \eta q_{\mathrm{H}}\,\Delta P - \int_0^{q_{\mathrm{H}}-q_{\mathrm{S}}} \eta\lambda_{\mathrm{F}} q\,\mathrm{d}q - \eta\lambda_{\mathrm{F}}(q_{\mathrm{H}} - q_{\mathrm{S}})q_{\mathrm{S}} - \int_0^{q_{\mathrm{S}}} P_q\,\mathrm{d}q \tag{44}$$

[16] HA for hybrid with anonymous floor.

Taking the derivative of G^{HA} with respect to the total amount of trading, q_{H}, we find that the total trading still follows Lemma 1

$$\Delta P = \lambda_{\mathrm{F}} q_{\mathrm{H}} \tag{45}$$

Thus, the overall trading efficiency of the market is again determined by the depth of the fast market. Of course, by offering what amounts to "cheaper" liquidity, the anonymous trading floor is more efficient in "stealing" trading from the fast market.

Again, having solved the problem of the total trading in this hybrid market, q_{H}, we find the condition that explains whether there will be any trading on the floor. As previously, taking the derivative of G^{HA} with respect to the amount of floor trading

$$\frac{\partial G^{\mathrm{HA}}}{\partial q_{\mathrm{S}}} = \eta \lambda_{\mathrm{F}} q_{\mathrm{S}} - (l_{\mathrm{o}} - l_{q_{\mathrm{S}}}) \tag{46}$$

we see that

$$\left[\frac{\partial G^{\mathrm{HA}}}{\partial q_{\mathrm{S}}}\right]_{q_{\mathrm{S}}=0} = 0 \tag{47}$$

and the trader will only trade at the floor when the second derivative at zero is positive.

Lemma 3. *In a hybrid market with anonymous trading floor, trading will be diverted from the fast market to the floor only if*

$$\lambda_{\mathrm{o}} < \lambda_{\mathrm{F}} \tag{48}$$

Thus, as suggested previously, we see that a trading floor can improve its competitive position vis-a-vis the fast market by offering anonymity.

5. Concluding remarks

After presenting a formal model of the liquidity and price discovery mechanisms in a hybrid market, that combines a trading floor with an electronic sweeping facility, the paper derives exact competitive criteria for fast and slow markets. Under the assumptions of the model here, a slow market only "steals" away liquidity from a fast market, and thus, hybrid markets do not generate more informed trading than stand-alone electronic markets, they only distribute the efficient amount of trading between the fast and the slow component.

It is shown that, when competing for liquidity with an electronic market $\lambda_{\mathrm{F}}^{-1}$ shares per dollar deep, a trading floor will be able to divert trades away from its electronic competitor only if it offers at least twice the depth of the electronic market. By stopping the front-running behavior of its specialists, a trading floor that provides anonymity to its traders is more attractive and will be able to compete with a fast market by only offering a marginally better depth, $\lambda_{\mathrm{o}}^{-1} > \lambda_{\mathrm{F}}^{-1}$.

Uncited references

Black (1995), Grosmann and Stiglitz (1980) and Harris (2003).

References

Back, K., & Baruch, S. (2004). Information in securities markets: Kyle Meets Glosten and Milgrom. *Econometrica*, *72*, 433–465.

Barndorff-Nielsen, O. E. (1998). Processes of Normal Inverse Gaussian type. *Finance and Stochastics*, *2*, 41–68.

Black, F. (1995). Equilibrium exchanges. *Financial Analysts Journal*, 23–29.

Glosten, L. (1994). Is the electronic open limit order book inevitable? *Journal of Finance*, *49*, 1127–1161.

Glosten, L., & Milgrom, P. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, *14*, 71–100.

Grosmann, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*, 393–408.

Harris, L. (2003). *Trading and exchanges: Market micro structure for practitioners*. Oxford, NY: Oxford University Press.

Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica*, *53*, 1315–1336.

Parlour, C., & Seppi, D. (2003). Liquidity-based competition for order flow. *Review of Financial Studies*, *16*, 301–343.

Peterffy, T., & Battan, D. (2004). Why some dealers and exchanges have been slow to automate. *Financial Analysts Journal*, 15–22.

Polimenis, V. (2005). A realistic model of market liquidity and depth. *Journal of Futures Markets*.

Santos, T., & Scheinkman, J. (2001). Competition among exchanges. *Quarterly Journal of Economics*, *116*, 1027–1061.

Seppi, D. (1997). Liquidity provision with limit orders and a strategic specialist. *Review of Financial Studies*, *10*, 103–150.

Viswanathan, S., & Wang, J. J. (2002). Market architecture: Limit order books versus dealership markets. *Journal of Financial Markets*, *5*, 127–167.