

Dual-Capacity Trading and the Quality of the Market*

AILSÄ RÖELL

London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom

This paper considers a securities market in which orders are channeled through professional broker-dealers such as London's market makers or the large banks operating on continental exchanges. If these dual-capacity dealers can judge the motives behind their customers' orders, they can trade profitably on their own account (even if they cannot "front run," that is, trade on their own account before executing a customer order). It is shown that the dealers have an incentive to satisfy roughly half of their customers' orders from their own inventory if they are sure that orders are liquidity-motivated and not based on inside information. As a result of dual-capacity dealing, transaction costs for liquidity-motivated traders in the aggregate fall, but they rise for those traders who are unable to convince any dealer that they have no inside information. The liquidity of the main market worsens, even though its effective liquidity for customers whose orders are partly filled from broker-dealer inventories improves. *Journal of Economic Literature* Classification Numbers: 020, 310, 520. © 1990 Academic Press, Inc.

1. INTRODUCTION

This paper considers the role of dual-capacity traders in an auction market for securities. Dual-capacity traders act both as brokers who bring clients' orders to the market and as dealers who trade on their own account. Examples that come to mind are London's market makers after the "Big Bang" of 1986, Chicago's floor traders, and the major banks in continental Europe.

This paper does not focus on the obvious conflict of interest inherent in determining a price when a dealer can execute clients' orders against his

* The author thanks Anat Admati, Jurgen Dennert, Marco Pagano, Pete Kyle, the participants in the 6th European Meeting on the Economics of Information, and the anonymous referees of this journal for many helpful comments and discussions. The support of the Center for Economic Research at Tilburg University is gratefully acknowledged.

own (or an associate's) book. Most exchanges have rules safeguarding against abuses, although the recent scandals in Chicago suggest that they are far from watertight. Instead, it will be taken for granted that dealers keep to rules designed to ensure fair pricing. In London, for example, market makers may satisfy brokerage orders in-house provided that the "best execution" is obtained: the price must be the best one quoted in the market. Similarly, in Italy banks that fill customer orders from their own inventory must do so at the market price reigning in the first following stock exchange batch auction in Milan.

Even so, in Italy it is commonly argued that banks are able to manipulate prices and take advantage of ordinary traders and that this may be a principal cause of the poor liquidity of the main market. Along similar lines, in London it has been argued that British market makers with a strong customer base have an unfair advantage over their (American) rivals and that the current stock exchange rule changes (which reduce the visibility of true available prices and recent trading history,¹ turning back the clock on some of the changes introduced at the time of the Big Bang) exacerbate this problem.

In this paper I attempt to model dual-capacity trading in a market where dealers are risk neutral and competitive, and orders are placed both by uninformed agents trading for liquidity purposes and by traders with some measure of inside information. Dual-capacity traders' competitive advantage rests in their ability to identify at least some of their brokerage customers as liquidity traders and to use this information in taking profitable positions.²

2. A MODEL WHERE DEALERS KNOW THEIR BROKERAGE CUSTOMERS

The setting investigated in this paper is one where broker-dealers are able to identify the customer who places an order with them and judge his motives for wishing to trade. This might stem from a long-standing relationship with the customer or a detailed knowledge of his current financial needs, so that the dealer can infer with some degree of certainty that his customer's wish to buy or sell does not stem from inside information.

For simplicity, it is assumed that the dealer either knows for sure that a particular customer is an uninformed liquidity trader or knows nothing at

¹ Market makers are now allowed to execute small orders in-house at the best quoted market price even if they themselves are not quoting that price on-screen. Also, the publication of large trades on-screen is to be delayed.

² This paper does not consider a second potential source of profit for broker-dealers: "front running," that is, trading on their own account after receiving customer orders but before executing those orders.

all about him (in practice, of course, intermediate degrees of knowledge are likely to prevail). Moreover, any given uninformed customer has at most one, if any, broker-dealer available who knows that he is uninformed. (This assumption is relaxed in Section 4, where competition among broker-dealers is examined.)

Suppose that aggregate liquidity traders' demand u is a mean-zero normally distributed variable with variance σ_u^2 . There are N risk-neutral broker-dealers, each of whom has a customer base which allows him to identify a portion of uninformed traders' demand u_i that has variance σ_i^2 , for $i = 1, \dots, N$. For ease of computation it is assumed that all dealers have an equally large (identifiably uninformed) customer base, so that $\sigma_i^2 = \sigma^2$ for all i . Individual traders' demands are independent, so that

$$\sigma_0^2 + \sigma_1^2 + \dots + \sigma_N^2 = \sigma_u^2,$$

where the index $i = 0$ refers to that portion of liquidity trading demand which none of the broker-dealers can certify as such ($\sigma_0^2 = 0$ is not ruled out). Such demand might be channeled through single-capacity (agency-only) brokers; or it might be placed via broker-dealers who are ignorant of the identity and motives of the agents placing the orders.

The broker-dealers submit net demand schedules on their own account of $Z_i(u_i, p)$, $i = 1, \dots, N$.

For simplicity let there be just one risk-neutral informed trader. This insider observes the best estimate of the security's value v , while all other market participants have a prior on v distributed normally with mean v_0 and variance V .³

It is assumed that there is a large enough number of competing market makers or "uninformed speculators" who submit price-quantity schedules to ensure that in equilibrium, the market price is equal to the expected value of the security, given aggregate net demand and public information.⁴ Total market demand (y) submitted to this group of competitive single-capacity market makers comes from the three main groups of agents identified above, the insider(s), the noise traders, and the noise traders' broker-dealers:

³ Note that we need not exclude any further uncertainty about the security's value (say, its liquidation value is $v + \varepsilon$, where ε is a further random variable whose conditional value given v is zero). Such additional uncertainty does not affect the behavior of the agents in our model, who are all presumed to be risk-neutral.

⁴ Kyle (1989) shows that in a model with M uninformed imperfectly competitive constant risk-averse market makers, there is a linear equilibrium in which each one provides a supply proportional to $(p - v_0)$. The efficient pricing assumed in our model emerges as the limit as $M \rightarrow \infty$ (so that both incentives for imperfectly competitive behavior and total market maker sector risk aversion vanish). Our assumption of competitive risk-neutral market making is a matter of convenience that does not drive the basic qualitative results obtained.

$$y = \sum_{i=0}^N u_i + \sum_{i=1}^N Z_i(u_i, p) + X(v, p).$$

Only y , but not its separate components, is observable to the market makers, or, more precisely, revealed by the market price in equilibrium. Perfect competition then ensures that the market is efficient in the sense that

$$p = E[v|y].$$

An *equilibrium* consists of demand schedules $Z_i(u_i, p)$ and $X(v, p)$ for the broker-dealers $i = 1, \dots, N$ and the insider, respectively, such that each of these agents maximizes expected profit, realizing that his trading affects the market price; and a price function $P(y)$ formulated by a competitive risk-neutral market making sector.

PROPOSITION 1. *The unique linear equilibrium of the model described above is given by*

$$P(y) = v_0 + \lambda y$$

$$X(v, p) = \beta(v - p)$$

$$Z_i(u_i, p) = -\delta u_i - \gamma(p - v_0), \quad i = 1, \dots, N,$$

where δ is the unique real root (for $\sigma_0^2 \neq 0$) of

$$(1 - \delta)^3 (2N - 1) \sigma^2 - (1 - \delta)^2 (N - 1) \sigma^2 + 2(1 - \delta) \sigma_0^2 - \sigma_0^2 = 0.$$

When $\sigma_0^2 = 0$, there are two roots, $\delta = N/(2N - 1)$ and $\delta = 1$. Note that $\frac{1}{2} \leq \delta \leq 1$ and $\delta \rightarrow \frac{1}{2}$ as $N \rightarrow \infty$. Given $\delta < 1$, the other parameters follow from

$$\beta = \sqrt{\frac{\sigma_0^2 + N(1 - \delta)^2 \sigma^2}{V}}$$

$$\gamma = \frac{1 - 2\delta}{1 - \delta} \beta$$

$$\lambda = \frac{1}{\beta} \frac{1 - \delta}{1 - \delta - N(1 - 2\delta)}.$$

Proof. See Appendix.

Observe that in equilibrium, broker-dealers supply about half⁵ of their identifiably uninformed customers' orders from their own inventory: in the proof of Proposition 1, it is shown that $\frac{1}{2} \leq \delta \leq N/(2N - 1)$. In effect, the broker-dealers act rather like ordinary monopolists facing a linear demand curve. Suppose a broker-dealer receives a buy order from a customer whom he knows is a liquidity trader. He knows the order conveys no information about the security's value. However, if the order is passed through to the market in full, it will drive up the market price; the market at large does not know whether it is liquidity-based or not, and therefore revises upward its estimate of the security's value. The broker-dealer takes advantage of this discrepancy between the market price and his own better estimate of value by satisfying part of the customer's order from his own inventory. Since the broker-dealer realizes that his sale will push the market price back down, he does not satisfy the customer's order in full. If he did, the net order brought to the market from himself and his customer would be zero, the market price would not rise above his estimate of the security's best value, and he would make zero expected profit. Instead, rather like an ordinary monopolist who supplies half the quantity that drives the price down to marginal cost, he fills about half the order.

Note also that $\gamma < 0$: broker-dealers' demand on their own account increases with the market price. When main market demand y increases, the uninformed market makers revise their estimate of v , and p , upward because they believe that the change may be due to trading by the insider as well as to liquidity trading. Each broker-dealer, however, is able to check that the increase in y is not caused by his own liquidity customers. He therefore attaches a relatively higher probability to the possibility that insider trading is responsible for the change in y . He therefore revises upward his estimate of v by a little more than the market makers do, and, accordingly, buys more of the security.

The somewhat unrealistic case where $\sigma_0^2 = 0$ deserves special mention. Here the broker-dealers as a group are able to identify the entire liquidity trading demand u . In that case there is always the following somewhat degenerate equilibrium. The insider formulates an infinitely elastic demand-supply schedule at price $p = v$. The market making sector provides no liquidity whatsoever ($\lambda = \infty$). The broker-dealers each satisfy, inelastically, their liquidity customers' demand ($Z_i = -u_i$, i.e., $\delta = 1$, $\gamma = 0$) and

⁵ In a simpler model in which dual-capacity traders are presumed to submit market orders or price-inelastic net demands (i.e., $\gamma = 0$), we obtain $\delta = \frac{1}{2}$ exactly, $\beta = \sqrt{(\sigma_0^2 + \frac{1}{4}N\sigma^2)/V}$, and $\lambda = 1/\beta$. Because dual-capacity traders are professional speculators who play a central role in the market, it seems more reasonable to model them as adjusting their demand to market conditions by setting price-quantity schedules.

provide no further liquidity. It is readily verified that the insider, the broker-dealers, and the market makers are unable to obtain positive profits in this equilibrium. Transaction costs are zero. The main market provides no liquidity; all liquidity demand is satisfied by the broker-dealers.

When $N = 1$, this is the only equilibrium for $\sigma_0^2 = 0$. When $N > 1$, there is in addition an equilibrium at which the market price does not perfectly reveal aggregate noise trading demand $u = u_1 + \dots + u_N$. Here $\delta = N/(2N - 1)$ and the insider and broker-dealers make positive profits, so that transaction costs for liquidity traders are nonzero.

At this point it seems appropriate to make some comments justifying our assumptions concerning the nature of the information exploited by the dual-capacity dealers.

Why need we assume that dealers know something about the identity of the traders who place the order flow? In our static model, where front running by dealers is excluded, the anonymous order flow to each dealer will not in itself convey any useful information over and above the aggregate order flow. If the main market and dealers $1 \dots N$ were to receive independent anonymous components of the uninformed order flow, $u_0 = u_{00} + \dots + u_{0N}$, with standard deviations $s_{00}, \dots, s_{0N}$, respectively, then it can be shown that in equilibrium a profit-maximizing insider would divide his total market order among the dealers in proportion to these standard deviations. Individual dealers' order flow would then convey no information on the insider's total trade that is not expressed in the aggregate order flow (a sufficient statistic for the insider's trade), and hence in the competitive market price. Thus dual-capacity dealers would not be able to trade profitably, and would simply take on the status of uninformed speculators. In short, as far as demand from unidentified agents is concerned, it makes no difference whether it is submitted to the market directly or via a broker-dealer.

Why not have the dual-capacity dealers identify insider orders rather than liquidity trading orders, and thus deduce some exclusive information about security value from their order flow? If able to identify an insider order, the dealer would have an incentive to trade in the same direction. This competition would spoil the market for the insider. Hence insiders have every incentive to hide behind anonymous intermediaries and/or to place orders directly on the main market via single-capacity (agency-only) brokers. In contrast, identifiable liquidity traders benefit from their dealer's trading from their own inventory, because it reduces adverse pressure on the market price.

Throughout this paper, the broker-dealer who knows that his client is uninformed is also assumed to see the size of his client's total net demand. If not, the client can profit at the dealer's expense by behaving strategi-

cally. Suppose that his true net demand is u_c . He can place an order for a large multiple Hu_c ($H \gg 1$) of his desired trade with his broker-dealer and reverse the excessive demand by a direct order of $(1 - H)u_c$ elsewhere. If the broker-dealer supplies a proportion δ of the announced order, the net demand to the main market from the client and the broker-dealer together will be

$$Hu_c + (1 - H)u_c - \delta Hu_c = (1 - \delta H)u_c.$$

Clearly, this strategy yields a better execution price, the higher the announced trade Hu_c . The broker-dealer suffers and, in fact, loses money if $H > 1/\delta$ because the market price is driven down by the combined orders while he is selling from his own inventory. If the broker-dealer suspects that a client may cheat him in this way, he will not wish to trade on his own account, and the client's order flow will appear in our model as a component of the anonymous order flow u_0 rather than of u_i for $i \geq 1$.

Thus, for the effects studied in our model to emerge, the broker-dealer not only needs to know that some of his clients are uninformed but also needs to be able to gauge their total order flow. This could be because the broker-dealer is able to monitor the client's total order flow to some degree, through contacts with other market professionals. Or perhaps he knows enough about his client to judge his trading needs. Or perhaps he simply trusts him. Last, observe that the relationship between the broker-dealer and a trusted client is of mutual benefit. By "cheating," the client risks his reputation and hence the recurring benefits of a long-term relationship.

3. COMPARISON OF MARKET EQUILIBRIUM WITH AND WITHOUT DUAL-CAPACITY DEALING

Let us compare the results obtained in Proposition 1 with the situation in which dual-capacity trading is banned. In that case, our model coincides with a limiting case discussed in Section 8 of Kyle (1989). Setting $\delta = \gamma = 0$ in Eq. (1) and (4) of the proof, we readily obtain

$$\beta^* = \sqrt{\frac{\sigma_0^2 + N\sigma^2}{V}}$$

$$\lambda^* = \frac{1}{\beta^*} = \sqrt{\frac{V}{\sigma_0^2 + N\sigma^2}}$$

Thus, in the absence of dual-capacity trading, the insider trades somewhat more vigorously (β is greater) in response to his information.

The presence of dual-capacity trading does harm the liquidity of the main market. Intuitively, dual-capacity traders offset a proportion $\delta \approx \frac{1}{2}$ of their liquidity customers' demand by supplying from their own inventory. This means that total liquidity trading on the main market is sparser. Then any order placed directly on the main market will have a greater impact on prices.

To see this, the price impact of an order from an anonymous liquidity trader (who does not have a relationship of trust with any broker-dealer) can be calculated. Solving the equilibrium conditions for the price in terms of exogenous variables, using Proposition 1, we obtain

$$\begin{aligned}
 p - v_0 &= \lambda \left(u_0 + \sum_{i=1}^N (1 - \delta) u_i + \beta(v - p) - N\gamma(p - v_0) \right) \\
 p - v_0 &= \frac{\lambda}{1 + \lambda(\beta + N\gamma)} \left(u_0 + \sum_{i=1}^N (1 - \delta) u_i + \beta(v - v_0) \right) \\
 &= \frac{1}{2\beta} \left(u_0 + \sum_{i=1}^N (1 - \delta) u_i \right) + \frac{1}{2} (v - v_0), \tag{1}
 \end{aligned}$$

i.e.,

$$\frac{dp}{du_0} = \frac{\lambda}{1 + \lambda(\beta + N\gamma)} = \frac{1}{2\beta}.$$

Hence the average transaction costs of the liquidity traders, $E[(p - v_0)u_0|u_0]$, are equal to $(1/2\beta)\sigma_0^2$.⁶

Similarly, in a market without dual-capacity trading

$$\frac{dp^*}{du} = \frac{1}{2\beta^*}$$

Since $\beta^* \geq \beta$, $dp/du_0 \geq dp^*/du$. Thus an order from a customer who cannot convince a dealer that he is uninformed moves the market price more than it would in the absence of dual-capacity trading. However, liquidity traders who are able to convince a broker-dealer that they have no information are better off. Their orders do not exert as much price pressure because the broker-dealer will satisfy roughly one-half of their order from his own inventory:

⁶ With linear schedules, $E[p - v_0|u_i] = (dp/du_i)u_i$. Hence expected transaction costs are given by $(dp/du_i)\sigma_i^2$, for $i = 0, 1, \dots, N$.

$$\begin{aligned}
\frac{dp}{du_i} &= \frac{1}{2\beta} (1 - \delta) \\
&= \frac{1}{2} \sqrt{\frac{V}{(\sigma_0^2/(1 - \delta)^2) + N\sigma^2}} \\
&< \frac{1}{2} \sqrt{\frac{V}{(\sigma_0^2 + N\sigma^2)}} \\
&= \frac{dp^*}{du}.
\end{aligned}$$

Thus these agents trade on better terms than they would in the absence of dual-capacity trading. Their total transaction costs are $(1/2\beta)(1 - \delta)N\sigma^2$.

We also calculate the ex ante (before observing v or u_i) expected profits of the insider and the dual-capacity traders.

$$\begin{aligned}
\text{Insider profit} &= (v - p)x \\
&= \beta(v - p)^2 \\
&= \beta \left[\frac{1}{2}(v - v_0) - \frac{1}{2\beta} \left(u_0 + \sum_{i=1}^N (1 - \delta)u_i \right) \right]^2
\end{aligned}$$

using Eq. (1).

Ex ante expected insider profit

$$\begin{aligned}
&= \frac{1}{4} \beta \left(V + \frac{1}{\beta^2} (\sigma_0^2 + N(1 - \delta)^2\sigma^2) \right) \\
&= \frac{1}{2} \beta V
\end{aligned}$$

using the expression for β given in Proposition 1.

Broker-dealers' total profit

$$\begin{aligned}
&= (v - p) \left(\sum_{i=1}^N z_i \right) \\
&= \left[\frac{1}{2}(v - v_0) - \frac{1}{2\beta} \left(u_0 + \sum_{i=1}^N (1 - \delta)u_i \right) \right] \left[-\delta \sum_{i=1}^N u_i \right. \\
&\quad \left. - N\gamma \left(\frac{1}{2}(v - v_0) + \frac{1}{2\beta} \left(u_0 + \sum_{i=1}^N (1 - \delta)u_i \right) \right) \right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2} \left[(v - v_0) - \frac{1}{\beta} \left(u_0 + \sum_{i=1}^N (1 - \delta) u_i \right) \right] \left[-\frac{1}{2} N\gamma (v - v_0) \right. \\
&\quad \left. - \frac{N\gamma}{2\beta} u_0 - \left(\delta + (1 - \delta) \frac{N\gamma}{2\beta} \right) \sum_{i=1}^N u_i \right].
\end{aligned}$$

Ex ante expected broker-dealers' profit

$$\begin{aligned}
&= \frac{N\gamma}{4} \left(-V + \frac{\sigma_0^2 + (1 - \delta)^2 N\sigma^2}{\beta^2} \right) + \frac{1}{2\beta} (1 - \delta)\delta N\sigma^2 \\
&= \frac{(1 - \delta)\delta}{2\beta} N\sigma^2.
\end{aligned}$$

Adding up, total expected profits of insider and broker-dealers

$$= \frac{1}{2\beta} (\sigma_0^2 + (1 - \delta)N\sigma^2).$$

This expression is, of course, equal to the total transaction cost to liquidity traders of both kinds. To see that total transaction cost is smaller than it would be in the absence of dual capacity, observe that

$$\begin{aligned}
&\frac{\text{Transaction cost with dual capacity}}{\text{Transaction cost without dual capacity}} \\
&= \frac{(1/2\beta) (\sigma_0^2 + (1 - \delta)N\sigma^2)}{(1/2\beta^*) (\sigma_0^2 + N\sigma^2)} \quad (2) \\
&= \sqrt{\frac{(\sigma_0^2 + (1 - \delta)N\sigma^2)^2}{(\sigma_0^2 + (1 - \delta)^2 N\sigma^2)(\sigma_0^2 + N\sigma^2)}} < 1 \quad \text{if } \delta > 0.
\end{aligned}$$

Thus the profits of the dual-capacity traders are more than offset by the reduction in insider profits, and transaction costs to liquidity traders fall. Intuitively, the overall quality of the market improves because less of the order flow "noise" is brought to the main market, to be exploited by the insider who hides his trading behind the noise. Liquidity traders who are known to a broker-dealer see a proportion δ of their order satisfied at zero transaction cost from the broker-dealers' inventories. The remainder of their orders, together with the anonymous liquidity orders, are brought onto the main market where transaction costs, as in Kyle (1985), are roughly proportional to the standard deviation of the order flow.

The effects of greater concentration in broker-dealing can also be examined in our model. Holding constant $N\sigma^2$ and $\sigma_0^2 (>0)$, the respective total variances of the liquidity trading demand that is and is not identifiable as such by the broker-dealer sector, it can be shown that as N increases, total transaction costs of liquidity traders rise. (In expression (2), the denominator is equal to the numerator plus a term $\delta^2 N\sigma^2 \sigma_0^2$. Totally differentiating the cubic expression for δ given in Proposition 1 with respect to N and δ , holding $N\sigma^2$ constant, we find that δ is decreasing in N . Then it is readily shown that expression (2) is increasing in N when $N\sigma^2$ is held constant.) Intuitively, as their number N increases, individual broker-dealers' information concerning the composition of total market demand is less superior to that of the market making sector. Therefore, they speculate less aggressively: both δ and $N\gamma$, their combined responsiveness to the market price, fall in absolute value. Insider profit increases with N while broker-dealers' joint profit falls (or stays constant in the case $\sigma_0^2 = 0$, $\delta = 1$), but by less. Total transaction costs rise.⁷

Related work by Cripps (1989) and Fishman and Longstaff (1989) considers a many-period model where broker-dealers may inherit inside information about the security's value directly by observing the orders placed by *informed* clients. These models (for Cripps, just the model of Section 5) are immune to our critique of Section 2 above (that the informed client would want to remain anonymous to prevent the broker-dealer from driving up the price against him by his trading) because the broker-dealers are effectively constrained not to trade on the information until one period after the client's order has been filled. Dealers are able to profit from the information only because the insider naively places a one-shot trade instead of exploiting his information optimally by trading repeatedly. Thus the profits of the broker-dealers are, in effect, profits the insider could have captured for himself by a dynamic trading strategy of the type studied by Kyle (1985). Since the broker-dealer profits stemming from observing the insider's traders are not obtained at the expense of the insider (when the insider trades first), they *must* be offset by increased transaction costs. In the Fishman and Longstaff model, in addition, the broker-dealers are able to identify uninformed traders, thus to some extent reducing the scope for insider profits and hence transaction costs, as in our model. The total impact on liquidity traders' transaction costs then becomes ambiguous, depending on which of the two effects dominates.

⁷ Note, from footnote 5, that changes in N , holding $N\sigma^2$ constant, have no effect in a model where broker-dealers submit market orders. As they do not speculate in response to changes in the market price, the accuracy of their inferences with regard to variables other than their own customers' order flow is irrelevant. In effect the equilibrium of the model with broker-dealers placing only price-inelastic market orders can be regarded as the limiting case $N \rightarrow \infty$ of our main model.

We have shown that both the customer and the broker-dealer will benefit from the latter's knowledge of the former's liquidity-motivated trading demand. What is crucial is this information, rather than the client-broker relationship that gives rise to it. For example, suppose that the broker-dealer knows about the client's trade, but the latter channels his order through another broker-dealer who knows nothing about him. Then the original broker-dealer will still trade as described in our model, while the broker-dealer who executes the order has no incentive to trade because he has no information over and above that of the market at large. This means that, once the division of information about clients and their orders has been fixed (and our model takes it as exogenous), there is no scope for either the broker-dealer or the client to bargain for a larger share of their joint surplus. The threat of breaking the customer-agent relationship is ineffective as it does not alter the outcome.

However, in the longer term it is possible to influence the allocation of the information about the client, and there is room to negotiate the division of the gains. For example, there should be negotiation at the time of setting up long-term relationship which will eventually allow the broker-dealer to judge the client's trading motives and needs. In particular, if broker-dealing is competitive, one would expect the client to be able to extract a contract for inexpensive below-cost related services from the broker-dealer, in anticipation of the latter's future profits from the relationship. Our model does not endogenize information gathering and dissemination by broker-dealers and their clients. It should be viewed as a building block for such a wider framework.

In general, our model focuses on expected gains and losses to the agents modeled but does not address the distortionary effects of such changes in rewards. For example, the reduction in insider profits means that there is less of an incentive to gather information. This may reduce the informational efficiency of the market. Our model is not complete enough to address the question of whether the current situation provides over- or underinvestment in information gathering. Indeed, it also fails to consider the distortionary effect of high transaction costs: liquidity traders' demand is taken to be exogenous and price-inelastic.

4. THE ROLE OF COMPETITION AMONG DUAL-CAPACITY DEALERS

In the model analyzed so far, it has been assumed that each liquidity trader can find only one, if any, broker-dealer whom he can convince that he is uninformed.

What happens when more than one broker-dealer knows that a particular segment of the order flow is pure liquidity trading? We consider one

such setting in which there are N segments of liquidity demand, about each of which K different dual-capacity traders are informed. For simplicity it is assumed that each dual-capacity trader has only one piece of such information,⁸ so that there are in total NK dual-capacity traders with some information.

PROPOSITION 2. *When $\sigma_0^2 \neq 0$, the unique linear equilibrium of the model described above is given by*

$$\begin{aligned} P(y) &= v_0 + \lambda y \\ X(v, p) &= \beta(v - p) \\ Z_{ij}(u_i, p) &= -\delta u_i - \gamma(p - v_0), \quad i = 1, \dots, N, j = 1, \dots, K, \end{aligned}$$

where δ is the unique real root (for $\sigma_0^2 \neq 0$) of

$$\begin{aligned} ((K+1)N-1)(1-K\delta)^3\sigma^2 - (N-1)(1-K\delta)^2\sigma^2 + \\ (K+1)(1-K\delta)\sigma_0^2 - \sigma_0^2 = 0. \end{aligned}$$

Note that $1/(K+1) \leq \delta \leq N/((K+1)N-1) \leq 1/K$ and hence $\delta \downarrow 1/(K+1)$ for large N .⁹ The other parameters are

$$\begin{aligned} \beta &= \sqrt{\frac{(\sigma_0^2 + N(1-K\delta)^2\sigma^2)}{V}} \\ \gamma &= \beta \left(\frac{1 - (K+1)\delta}{1 - K\delta} \right) \\ \lambda &= \frac{1}{\beta} \frac{1 - K\delta}{1 - K\delta - KN(1 - (K+1)\delta)}. \end{aligned}$$

Proof. Not shown. Exactly analogous to the method used in proving Proposition 1. ■

⁸ In a model with price-inelastic dual-capacity trader demands, this assumption would have no effect on the outcome.

⁹ It can be shown that with price-inelastic dual-capacity trader demand, $\delta = 1/(1+K)$ as in the linear Cournot–Nash model. Then any uninformed trader identified as such by K dealers will see a proportion $K/(K+1)$ of his order satisfied from dealers' inventories, with only a proportion $1/(K+1)$ going to the main market. Hence,

$$\beta = \sqrt{\frac{\sigma_0^2 + \left(\frac{1}{K+1}\right)^2 \sigma^2}{V_0}}, \quad \lambda = \frac{1}{\beta}.$$

Observe that as K increases $(1 - K\delta) \approx 1/(K + 1)$ decreases. As the number of competing dealers who know that a particular trade is liquidity-motivated increases, the proportion not supplied directly from dealers' inventory decreases. Liquidity traders have an incentive to convince as many dealers as possible that the order that they plan to place is not motivated by inside information. "Sunshine trading," in which agents publicly announce in advance their intention to make a large deal, is an example discussed in further detail and in the context of a competitive model with risk-averse agents by Admati and Pfleiderer (1990) and also by Gennotte and Leland (1990). In practice such announcements often fail to achieve the desired effect for two reasons. First, market professionals may not believe that the proposed deal is not information-generated. Second, such an announcement may not be credible for reasons discussed at the end of Section 2: even bona fide liquidity traders have an incentive to announce a very large intended trade, and then place a smaller actual order. Dealers who trade based on the announcement then take losses.

In our model, the market price is more responsive to unidentified liquidity demand than to identified liquidity demand:

$$\frac{dp}{du_i} / \frac{dp}{du_0} = 1 - K\delta < 1 \quad (\rightarrow 0 \text{ as } K \rightarrow \infty).$$

This result mirrors the findings of Gennotte and Leland (1990) in a competitive model with risk-averse agents. In their model the market price also becomes substantially less responsive to the order flow when more agents are informed that it is liquidity-driven.

In the limit, as $K \rightarrow \infty$, a situation where only the anonymous component of the order flow determines market liquidity is achieved:

$$\beta \rightarrow \sqrt{\frac{\sigma_0^2}{V}}$$

$$\lambda \rightarrow \frac{1}{\beta}$$

$$(1 - \delta K) \rightarrow 0.$$

Here main market liquidity is minimal, but certified liquidity traders exert no pressure on prices and thus trade at near-zero transaction cost. Broker-dealers are too competitive to make profits, and insider profits are minimal. Thus aggregate transaction costs to liquidity traders as a group are minimized; however, they are borne entirely by those liquidity traders who are unable to convince the market professionals that they are uninformed.

In theory, much the same effect may be achieved indirectly if dual-capacity firms are forced to announce their price–quantity schedules publicly during the auction market tatonnement. The outcomes described in Propositions 1 and 2 would then no longer be an equilibrium because the schedules $Z_i(u_i, p)$ reveal $\{u_1, \dots, u_N\}$ publicly to all other market participants, who adjust their behavior accordingly. Instead, the limiting outcome described above would be an equilibrium.

Thus with dual-capacity firms forced to reveal their price–quantity schedules for proprietary trading, the equilibrium that fully reveals $u_i + \dots + u_N$ could emerge:

$$\begin{aligned} P(y) &= v_0 + \lambda \left(y - \sum_{i=1}^N u_i \right) \\ X(v, p) &= \beta(v - p) \\ Z_i(u_i, p) &= 0, \quad i = 1, \dots, N, \end{aligned}$$

where

$$\begin{aligned} \beta &= \sqrt{\frac{\sigma_0^2}{V}} \\ \lambda &= \sqrt{\frac{V}{\sigma_0^2}}. \end{aligned}$$

This equilibrium minimizes total transaction costs for liquidity traders.¹⁰

London's market makers are an example of a group of dual-capacity dealers who are to some degree forced to publicize their trading strategy in the form of electronically displayed on-screen bid and ask quotes. However, these quotes are not fully informative because they are binding only for small trades (for large trades, as suggested by our model, market makers prefer to negotiate a price that depends on the identity of their counterparty). Moreover, now dual-capacity firms are allowed to satisfy small orders from their own inventory at will at the best price quoted in the market even if they themselves are not quoting that price. Hence there

¹⁰ In this equilibrium the agents who have private information (the dual-capacity dealers) have no incentive to trade and thus express their information (the u_i) in the final equilibrium; see Hellwig (1980) for a discussion of this problem in the context of a competitive rational expectations equilibrium. Presumably the information is conveyed during the preceding tatonnement process, in which dealers have no incentive to hold back information (given that in the final equilibrium they will be held to zero trading on their own account anyway) but also no incentive to reveal it.

is now even more scope for adopting a proprietary trading policy that is not visible to other market participants via the bid-ask quotes. This should thus lead to higher transaction costs for liquidity traders in the aggregate.

5. POLICY IMPLICATIONS

In our model dual-capacity trading reduces total transaction costs for liquidity traders. True, the liquidity for anonymous orders on the main market deteriorates. But this effect is more than offset by benefits to customers of dual-capacity firms which partly fill orders from their own inventory, resulting in less price pressure and improved execution quality. A Chicago Mercantile Exchange Panel defended dual-capacity trading (*Wall Street Journal*, April 20, 1989) on these grounds: "Dual trading is needed . . . to maintain enough liquidity."

Our model identifies two potential disadvantages associated with dual-capacity trading. First, it may be considered unfair that some traders' transaction costs actually rise though the total falls. Equal treatment of all potential investors is a principle on which many current rules are based.¹¹ Second, insiders' profits are reduced. This might be undesirable if the market currently provides insufficient incentives for gathering information.

In Section 4 it was shown that these effects are particularly strong if liquidity traders can convince multiple competing broker-dealers that they have in mind a trade that is not information-driven, or if broker-dealers are forced to make public their price-quantity schedules.

One final word of caution: Our model does not explore the full range of effects associated with dual-capacity trading. In particular, our static approach precludes an analysis of "front running" whereby dual traders trade on their own account in advance of customer order execution. We also do not address one of the prime practical reasons why broker-dealers in continental Europe have diverted order flow away from the main exchange: fixed brokerage commissions. When these exceed order processing costs, there is a strong incentive to cross orders in-house or fill them from their own inventory. This particular motive for in-house execution should disappear once commission rates are left to be determined by competitive market forces.

¹¹ For example, the London and NASDAQ rule requiring market makers to quote "firm" prices (at which they are obliged to trade with all comers) for deals up to a certain size means that it is difficult to take advantage of inexperienced traders. But it also means that market makers are less able to protect themselves against known insiders and must therefore charge higher average spreads.

APPENDIX: PROOF OF PROPOSITION 1

We look for a Nash equilibrium in trading strategies. A justification for the method used is given in Section 5 of Kyle (1989).

I. The *insider* chooses a net demand schedule $X(v, p)$ that maximizes his expected profits, given that he knows v , the value of the security, and infers (from the market price) the noise traders' demand net of the resulting supply from broker-dealers:

$$\begin{aligned} \max_{\{x\}} (v - p) x, \quad \text{where } p = v_0 + \lambda(W - N\gamma(p - v_0) + x) \text{ and } W \\ \equiv \bar{u} - \delta \sum_{i=1}^N \bar{u}_i. \end{aligned}$$

Substituting in for p and taking the first-order conditions, we obtain

$$\begin{aligned} x &= \frac{1 + \lambda N\gamma}{2\lambda} (v - v_0) - \frac{1}{2} W \\ x &= \frac{1 + \lambda N\gamma}{\lambda} (v - p) \\ \beta &= \frac{1 + \lambda N\gamma}{\lambda}. \end{aligned} \tag{1}$$

II. *Broker-dealer* i maximizes expected profits, given u_i and the market price (from which he infers $K = \bar{u}_0 + \sum_{j \neq i} (1 - \delta)\bar{u}_j + \beta(\bar{v} - v_0)$):

$$\begin{aligned} \max_{\{z_i\}} E[(\bar{v} - p) z_i], \quad \text{where } p = v_0 + \lambda(u_i + K - \beta(p - v_0) \\ - \sum_{j \neq i} \gamma(p - v_0) + z_j). \end{aligned}$$

Substituting in for p , taking the first-order condition, and rewriting in terms of p , we obtain

$$z_i = \frac{1 + \lambda\beta + \lambda(N - 1)\gamma}{\lambda} (E[\bar{v} - v_0] - (p - v_0)).$$

But

$$\begin{aligned} E \left[\bar{v} - v_0 | \bar{u}_0 + \sum_{j \neq i} (1 - \delta)\bar{u}_j + \beta(\bar{v} - v_0) = K \right] \\ = \frac{\beta V}{\beta^2 V + \sigma_0^2 + (N - 1)(1 - \delta)^2 \sigma^2} K \end{aligned}$$

$$= \frac{\beta V}{\beta^2 V + \sigma_0^2 + (N-1)(1-\delta)^2 \sigma^2} \left[\frac{1 + \lambda\beta + (N-1)\gamma}{\lambda} (p - v_0) - u_i - z_i \right].$$

Hence

$$\begin{aligned} & \left[1 + \frac{\beta V}{\beta^2 V + \sigma_0^2 + (N-1)(1-\delta)^2 \sigma^2} \frac{1 + \lambda\beta + (N-1)\gamma}{\lambda} \right] z_i \\ &= \frac{1 + \lambda\beta + \lambda(N-1)\gamma}{\lambda} \left[\left(\frac{\beta V}{\beta^2 V + \sigma_0^2 + (N-1)(1-\delta)^2 \sigma^2} \right. \right. \\ & \quad \left. \left. \frac{1 + \lambda\beta + (N-1)\gamma}{\lambda} - 1 \right) (p - v_0) \right. \\ & \quad \left. - \left(\frac{\beta V}{\beta^2 V + \sigma_0^2 + (N-1)(1-\delta)^2 \sigma^2} \right) u_i \right] \\ \gamma &= \frac{1 + \lambda\beta + \lambda(N-1)\gamma}{\lambda} \\ & \quad \left(\frac{1 - \frac{1 + \lambda\beta + \lambda(N-1)\gamma}{\lambda} \frac{\beta V}{\beta^2 V + \sigma_0^2 + (N-1)(1-\delta)^2 \sigma^2}}{1 + \frac{1 + \lambda\beta + \lambda(N-1)\gamma}{\lambda} \frac{\beta V}{\beta^2 V + \sigma_0^2 + (N-1)(1-\delta)^2 \sigma^2}} \right) \quad (2) \end{aligned}$$

$$\delta = \frac{\frac{1 + \lambda\beta + \lambda(N-1)\gamma}{\lambda} \frac{\beta V}{\beta^2 V + \sigma_0^2 + (N-1)(1-\delta)^2 \sigma^2}}{1 + \frac{1 + \lambda\beta + \lambda(N-1)\gamma}{\lambda} \frac{\beta V}{\beta^2 V + \sigma_0^2 + (N-1)(1-\delta)^2 \sigma^2}}. \quad (3)$$

III. *Semistrong market efficiency.* A competitive group of market makers or speculators ensures that

$$\begin{aligned} p &= E[\bar{v}|y] \\ &= E \left[\bar{v}|y = \bar{u} + \beta(\bar{v} - p) - N\gamma(p - v_0) - \sum_{i=1}^N \delta \bar{u}_i \right] \\ p - v_0 &= \frac{\beta V}{\beta^2 V + \sigma_0^2 + N(1-\delta)^2 \sigma^2} (y + (\beta + N\gamma)(p - v_0)) \end{aligned}$$

$$p = v_0 + \frac{\beta V}{\sigma_0^2 + N(1 - \delta)^2 \sigma^2 - \beta V N \gamma} y$$

$$\lambda = \frac{\beta V}{\sigma_0^2 + N(1 - \delta)^2 \sigma^2 - \beta V N \gamma}. \quad (4)$$

IV. *Solution of equation system (1) to (4).* Inserting (4) into (1), we have

$$\beta = \frac{\sigma_0^2 + N(1 - \delta)^2 \sigma^2 - \beta V N \gamma}{\beta V} + N \gamma$$

$$\beta^2 V = \sigma_0^2 + N(1 - \delta)^2 \sigma^2, \quad (5)$$

and using (1) and (3) in (2),

$$\gamma = (2\beta - \gamma)(1 - 2\delta).$$

Hence

$$\gamma = \beta \frac{(1 - 2\delta)}{(1 - \delta)} \quad (6)$$

and from (3), using (1),

$$1 - \delta = \frac{1}{1 + (2\beta - \gamma) \frac{\beta V}{2\sigma_0^2 + (2N - 1)(1 - \delta)^2 \sigma^2}}.$$

Using (6) and (5) to eliminate γ and β in this equation, we obtain the cubic expression for δ :

$$(1 - \delta)^3 (2N - 1) \sigma^2 - (1 - \delta)^2 (N - 1) \sigma^2 + 2(1 - \delta) \sigma_0^2 - \sigma_0^2 = 0.$$

Since the sum of the first two terms must be opposite in sign to the sum of the last two terms, it is readily verified that the cubic equation's real roots must lie in the interval $[\frac{1}{2}, 1]$; indeed, more precisely, $\delta \in [\frac{1}{2}, N/(2N - 1)]$. Inside that interval, the slope with respect to $(1 - \delta)$,

$$3(1 - \delta)^2 (2N - 1) \sigma^2 - 2(1 - \delta) (N - 1) \sigma^2 + 2\sigma_0^2,$$

must be positive whenever the cubic expression equals zero, since that event $(1 - \delta)^3 (2N - 1) \sigma^2 - (1 - \delta)^2 (N - 1) \sigma^2 > 0$ (given that $2(1 - \delta) \sigma_0^2 - \sigma_0^2 < 0$). Hence, there is only one real root when $\sigma_0 \neq 0$. The expressions for β , λ , and γ follow directly. ■

REFERENCES

- ADMATI, A. R., AND PFLEIDERER, P. (1990). Sunshine trading: The effects of preannouncement on traders' welfare and on price variability, unpublished, Graduate School of Business, Stanford University, revised version.
- CRIPPS, M. (1989). Dealer behaviour and price volatility in asset markets, unpublished, Department of Economics, Warwick University.
- FISHMAN, M. J., AND LONGSTAFF, F. A. (1989). Dual trading in futures markets, unpublished, Kellogg Graduate School of Management, Northwestern University, November.
- GENNOTTE, G., AND LELAND, H. E. (1990). Market liquidity, hedging and crashes, *Amer. Econ. Rev.* **80**, 999–1021.
- HELLWIG, M. F. (1980), On the aggregation of information in competitive markets, *J. Econ. Theory* **22**, 477–498.
- KYLE, A. S. (1985). Continuous auctions and insider trading, *Econometrica* **53**, 1315–1335.
- KYLE, A. S. (1989). Informed speculation with imperfect competition, *Rev. Econ. Stud.* **56**, 317–355.