Structural Organization of Secondary Markets: Clearing Frequency, Dealer Activity and Liquidity Risk

Author(s): Kenneth D. Garbade and William L. Silber

Source: *The Journal of Finance*, Jun., 1979, Vol. 34, No. 3 (Jun., 1979), pp. 577-593

Published by: Wiley for the American Finance Association

Stable URL: https://www.jstor.org/stable/2327427

THE JOURNAL OF FINANCE • VOL XXXIV, NO. 3 • JUNE 1979

The Journal of FINANCE

VOL. XXXIV JUNE 1979 No. 3

Structural Organization of Secondary Markets: Clearing Frequency, Dealer Activity and Liquidity Risk

KENNETH D. GARBADE and WILLIAM L. SILBER*

1. Introduction

A FINANCIAL INSTRUMENT IS commonly considered liquid if it has at least one of two attributes. First, the instrument may be traded in a market with a sufficient number of participants to make feasible purchases and sales on short notice at prices near the contemporaneous equilibrium value of the instrument. Thus, common stocks actively traded on the New York Stock Exchange are considered more liquid than most municipal bonds. There exists an almost purely competitive market for the former securities, whereas secondary market sales of municipals often require substantial price discounts if they are to be completed quickly. Second, an asset is regarded as liquid if its equilibrium value is unlikely to change substantially over a given interval of time. Even though short-term municipals do not trade in an active secondary market, their values are not as volatile as those of common stock issues. A seller of a short-term municipal can therefore spend time searching for a favorable trading partner without bearing excessive price risk during the search process.

In this paper we show that the two aspects of liquidity just noted, the number of market participants and equilibrium price volatility, are important determinants of the structural characteristics of secondary markets. The key structural characteristics are the frequency of market clearing and the level and effect of dealer participation. The relationships between these characteristics and their determinants are derived using a concept of market performance which focuses on liquidity risk. This measure of risk is defined as the variance of the difference between the equilibrium value of an asset at the time a market participant decides to trade and the transaction price ultimately realized. We also show that this measure of liquidity risk encompasses both aspects of liquidity noted above.

Clearing Frequency of a Market

Secondary markets differ with respect to clearing frequency. For example, the New York Stock Exchange operated as a twice-daily call auction before the Civil

* The authors are Associate Professor and Professor, respectively, of Economics and Finance at the Graduate School of Business Administration, New York University. We are indebted to William J. Baumol whose comments sparked the development of this paper. Financial Support from NSF Grant #RDA75-17953 is gratefully acknowledged.

War. Modern examples of periodic call markets include the Paris Bourse, Eiteman and Eiteman, [3] and the Tel Aviv Stock Exchange, Silber, [15]. At the other extreme, since the Civil War clearings on the New York Stock Exchange can be considered continuous. The liquidity risk of trading a particular financial asset depends on the frequency of market clearings. To see this effect, consider that our price variance measure may be separated conceptually into two components. First is the variance of the change in the equilibrium value of an asset between the time an investor decides to trade and the time the trade is completed. This component of risk is clearly an increasing function of the time taken to complete a trade. Its presence creates a demand for prompt order execution, i.e., frequent or continuous clearing.

The second part of liquidity risk is the variance of the difference between contemporaneous transactions prices and equilibrium values. Transactions prices, i.e., market clearing prices, can differ from equilibrium prices when potential participants in a market do not enter orders. Except for the timeless and spatially integrated world of a Walrasian auction, market clearing prices are always determined by subsets of transactors. Thus, market clearing prices will, in general, differ from the equilibrium price derived from a Walrasian auction as a consequence of the temporal fragmentation of potential market participants. Ceteris paribus, the longer the time between clearings, the greater the number of participants in a given clearing. Thus, transactions prices in markets with frequent or continuous clearings are more likely to exhibit transient random deviations from current equilibrium values than are prices in markets with less frequent and more consolidated clearings.[1] To reduce the deviations of transactions prices from equilibrium values one would prefer to execute orders in consolidated markets with infrequent clearings.

The optimal clearing frequency of a market is that time interval which minimizes total liquidity risk, i.e., the sum of the variance of equilibrium price changes (an increasing function of the time interval) and the variance of transient random differences between transactions prices and equilibrium values (a decreasing function of the time interval). We expect that observed market structures reflect the optimal clearing frequencies of different securities. This follows from the assumption that investors will trade in those markets organized to minimize their exposure to liquidity risk.

In section 3 we develop a model of the behavior of securities prices in a public market, and then consider the implications of that model for market organization. In particular, we show that the optimal frequency of market clearings is an increasing function of the equilibrium price volatility of a security and an increasing function of the number of market participants. The role of technological innovations in expanding the size of a market is also considered.

[1] The Walrasian auction assumes that all potential transactors have instantaneous access to the market whenever the market clears, i.e., whenever transactions are executed. In this paper we consider the consequences of clearings with incomplete investor participation. The New York Stock Exchange noted [13, p. 1] that, "Dispersion of orders over time can result in temporary imbalances of supply and demand unrelated to real changes in values." In a companion paper we expect to demonstrate that dispersion of orders over multiple trading centers, i.e., spatial fragmentation, can be analyzed in a fashion similar to that presented here.

Dealer Participation

Another important characteristic of a market is the presence of professional securities dealers who trade against transient imbalances in public purchase and sale orders. We show that dealer participation in a market increases the optimal frequency of clearing or, looked at another way, reduces the length of time typically required to complete a transaction. We also show that dealer participation reduces the liquidity risk born by public transactors. Thus, investors may be expected to prefer to trade in dealer markets rather than what may be defined as purely public markets.[2] We suggest that dealers are more active the larger the size of the market and the lower the volatility of equilibrium prices. Thus, these two aspects of a financial instrument affect the frequency of market clearings and the level of liquidity risk both directly and indirectly through dealer activity.

In section 4 we develop a model of how dealers might behave in the market introduced in section 3, and we then present the implications of their behavior for optimal clearing frequencies (increased) and for the magnitude of the liquidity risk borne by public transactors (decreased). Before proceeding with the formal analysis, the next section discusses a number of historical and contemporaneous examples of alternative forms of market organization. These illustrations should motivate the broad conclusions which flow from the analysis of sections 3 and 4.

II. Comparative Market Organization

The analytical results derived below suggest that as a market becomes larger, or as securities prices become more volatile, investors will be led to prefer exchange mechanisms which provide for faster execution of their orders. In our terms this means more frequent market clearings, approximating, in the limit, continuous trading. The institutional evolution of the New York Stock Exchange (NYSE) illustrates clearly these principles.

The Exchange traces its ancestry to a meeting of securities brokers held in May, 1792 to formalize trading practices in the securities markets of New York City. The brokers organized their market as a periodic auction, usually with a morning session and an afternoon session. Securities were called for trading sequentially, and buyers and sellers competed collectively to get the best price on their orders. The basic structure of periodic auctions on the New York Stock and Exchange Board (as the market was known after March, 1817) lasted until the Civil War.

With the advent of the Civil War, securities prices became measurably more volatile as a consequence of the flow of war-related information. As our model suggests, this increased volatility led to a demand for opportunities for more frequent trading, Sobel, [16]. When the governors of the New York Board refused to introduce a continuous market during the war, trading largely moved onto off-board markets. Sobel [16] notes that, "there were literally dozens of separate trading groups in operation by 1865. . . ." The most important of these was the Open Board of Stock Brokers, organized in 1864, with continuous trading between the hours of 8:30 AM and 5:00 PM. By the middle of 1865 the volume of

[2] It is interesting to observe that the the Securities and Exchange Commission considered, in the 1930s, the feasibility of prohibiting dealer activity on the New York Stock Exchange. See Sobel [17, pp. 22–23]. Our analysis suggests the initiative was misguided.

transactions on the NYSE (renamed in 1863) was only a tenth the volume on the Open Board, Sobel, [16]. The demand for trading facilities during the Civil War was so great that there even existed an Evening Exchange, so that purchases and sales of securities could be completed in New York City at any time of the day or night, Sobel, [16].

Following the Civil War the Open Board and the NYSE continued with their respective trading practices. However, in 1868 and 1869 the governors of the two markets negotiated a merger which was ultimately concluded on May 8, 1869. The consolidated NYSE employed both discrete call auctions and continuous trading, but by 1870 the call auctions had become relatively unimportant (they were formally eliminated for stocks in 1882).

Securities prices became relatively stable with the end of the Civil War, yet continuous trading survived and in fact flourished as the main form of secondary market organization. The explanation is found in a series of technological innovations which expanded the number of participants with ready access to the market. The first innovation was the stock ticker, invented in 1867 by E. A. Calahan and put into operation at the end of that year by the Gold and Stock Telegraph Company, Hotchkiss, [9]. The ticker was important because it permitted market participants located away from the Exchange floor to follow even a continuous market with virtually no time delay. The ticker fostered continuous trading because it created a larger number of fully informed market participants.

Although tickers provided price information quickly to participants located away from the floor of the NYSE, the devices had no capability for transmitting orders back to the floor. Orders continued to be sent via general telegraph companies, and suffered the attending delays between origination and execution. These delays were largely eliminated with the introduction of private brokers wires. The first such wire opened in 1873 between the uptown head office of a New York securities firm and its downtown office in the financial district, Meeker, [11]. This private wire (actually, a telegraph line leased for private use from a general telegraph company and operated by the leasee) facilitated the transmission of orders to the floor of the Exchange by eliminating the necessity to deliver a telegraph message to an uptown office and to redeliver the message from a downtown receiving office. In subsequent years brokers wires were extended to firms with offices outside New York City.

The introduction of the stock ticker and brokers wires expanded the size of the market for NYSE-listed issues in the sense of increasing the number of participants with rapid access to the floor. Meeker [11] noted the expanded scope of the New York markets made possible by technology when he observed that, "..... the Stock Exchange system [including wire houses and tickers] has practically annihilated the considerations of time and space in the operation of America's principal securities market".[3] Our model suggests that the increased size of the NYSE markets was a factor contributing to the dominance of continuous trading even in the relatively stable post-Civil War environment.

The organization of the municipal debt markets provides a contemporary example of the effect of equilibrium price volatility and market size on the

[3] Garbade and Silber [7] provide a more detailed evaluation of the impact of telegraphic communication on financial market integration. Garbade [4] evaluates similar developments in the government securities market during the 1970s.

organizational structure of a market. As debt instruments bearing (outside of exceptional cases) little credit risk, municipals exhibit relatively low equilibrium price volatility over intervals of time up to a few days. There are, however, an enormous number of municipal issues outstanding, many in issue sizes no greater than a few million dollars. The number of market participants actively seeking to buy or sell a given issue at a particular point in time is quite small. The low price volatility and limited market size characteristic of municipals leads us to expect that the time required to complete a favorable transaction is relatively long or, in our terms, that the municipal market clears relatively infrequently. This is in fact the case. Institutional traders of municipal bonds (bank trust departments) rely on bond brokers to locate potentially compatible trading partners. The traders call their purchase and sale interests into as many as forty bond brokers in the morning and then wait until they receive replies in the afternoon before they choose the best bids and offerings. Their behavior appears very much like a process of once-daily market clearings.

III. A Model of a Public Market

The historical experience reported in the preceeding section supports the intuitive analysis of liquidity risk and market organization discussed in section 1. We now present a formal model that can be used to derive more precise relationships between liquidity risk and market organization. Our model assumes an exchange structure characterized by periodic clearings, such as the New York Stock Exchange before 1869 or the contemporary gold fixings market in London, Jarecki, [10]. The time between clearings is an unspecified variable whose optimal value is to be determined. Markets in which securities transactions are completed quickly, such as the present day NYSE, can be considered a limiting case where the time between clearings becomes small, approximating continuous trading in the limit. To focus attention on the issues of immediate interest, we consider here a market in which there are only public transactors. The analysis of a market in which dealers also participate is more complex and is deferred to section 4.

The process by which public orders enter the market is simply specified. Suppose there are K public transactors participating in a given clearing, each with an endowment of E of the security. Each transactor tenders his full endowment to the market. This behavior is justified by the assumption that the market is perfectly competitive, so that no transactor perceives his activities as affecting the market clearing price. The gross demand schedule of the i^{th} transactor as a function of the market clearing price p is given by:

$$D_i(p) = E + a(r_i - p) \qquad a > 0 \tag{1}$$

r_i is the reservation price of the i^{th} participant. If the market clears at a price greater than r_i that transactor will be a net seller of securities, so that $D_i < E$. If it clears at a price below r_i he will be a buyer. For simplicity we assume the values of the endowment E and the slope a of the demand schedule are common to all transactors.

Since there are only public transactors in the present model the market will clear at the price which equates aggregate public supply with aggregate demand, or when:

$$KE = \sum_{i=1}^{K} [E + a(r_i - p)] \tag{2}$$

The market clearing price thus equals the mean reservation price of the public participants. Let r denote that market clearing price:

$$r = K^{-1} \sum_{i=1}^{K} r_i \tag{3}$$

The model is completed by specifying how investor reservation prices are related to each other in the same market meeting, and through time in different meetings. We assume there exists, at time t, an unobservable equilibrium value m_t for the security, which is not necessarily equal to the market clearing price. The reservation price of a given transactor at time t is a normally distributed random variable with mean m_t and variance σ^2. The difference between m_t and the reservation prices of different investors are uncorrelated. Let r_t denote the mean reservation price at time t which, according to the discussion preceeding equation (3), means that r_t is the market-clearing or transactions price at that time. Equation (3) implies the distribution of r_t is:

$$r_t = m_t + f_t \tag{4a}$$

$$f_t \sim N(0, \sigma^2/K) \tag{4b}$$

For simplicity we assume f_t is a serially uncorrelated process.

As K grows large the distribution of r_t collapses towards a singular distribution at m_t. It is therefore natural to speak of m_t as the equilibrium value of the asset, since the market clearing price r_t approaches m_t in probability as the number of participants in the market at time t grows large. Our definition of equilibrium value agrees with its usual definition as the market clearing price when the number of participants active in the market is large.

The equilibrium value of the asset is assumed to evolve continuously through time in a random walk. The change in equilibrium value from one market clearing to the next therefore follows the process:

$$m_t = m_{t-1} + e_t \tag{5a}$$

$$e_t \sim N(0, \tau\psi^2) \tag{5b}$$

ψ^2 is the variance of the change in equilibrium value per unit time. τ is the time between market clearings. e_t is a serially uncorrelated process and is uncorrelated with the random process f_t in equation (4a). One could make other assumptions on the dynamics of equilibrium value, including secular drift, without altering the results of interest in this paper.

We assume that over an interval of time of length τ there will be $w\tau$ transactors entering the market. w is the time rate of exposure of transactors to the market. In general, w will vary across securities as a function of the extent of trading interest in those securities. It will also depend on the sophistication of communication technologies.[4] We pointed out above that the innovations of stock tickers

[4] The extent of trading in a security should be determined, in part, endogenously, as in Garbade and Silber [6]. For our purposes, this is an unnecessary complication since we are interested primarily in the impact of exogenous changes in market size such as those due to technological innovations.

and private brokers wires expanded significantly the number of investors with ready access to the NYSE. This expansion is represented in the present model by a larger value of w.

If the time between market clearings is τ, there will be $K = w\tau$ participants in each clearing. Smaller values of τ imply smaller markets in each clearing. Although the foregoing model was formally developed for discrete values of K, the equations remain valid for the continuum case as well if we interpret K as a scale parameter characterizing the size of a clearing. From equation (4) we have that the transient difference between the clearing price r_t and the contemporaneous equilibrium value m_t is:

$$r_t = m_t + f_t \tag{6a}$$

$$f_t \sim N(0, \sigma^2/w\tau) \tag{6b}$$

The change in equilibrium value between clearings is given by equation (5).

Liquidity Risk and the Optimal Clearing Frequency of a Public Market

Using the model of equations (5) and (6) we can now derive an expression for liquidity risk as a function of the price volatility of the asset being traded and the size of its market. We then determine the market structure which minimizes this liquidity risk.

Liquidity risk is defined as the variance of the difference between the equilibrium value (m) of an asset at the time an investor decides to trade and the transaction price (r) ultimately realized on that trade. The present model assumes that the typical investor makes the decision to trade at time $t - ½$ for execution at time t. His liquidity risk is then:

$$\mathrm{Var}[r_t - m_{t-1/2}] = \mathrm{Var}[(r_t - m_t) + (m_t - m_{t-1/2})] \tag{7}$$

We can treat the random term $e_t = m_t - m_{t-1}$ in equation (5a) as the sum of two independent random terms, $m_t - m_{t-1/2}$ and $m_{t-1/2} - m_{t-1}$, each written with variance $½\ \tau\psi^2$. Neither of these terms is correlated with $r_t - m_t = f_t$. From equations (5) and (6), the liquidity risk V_p of an investor trading in a public market is then:

$$V_p = \sigma^2/w\tau + ½\ \tau\psi^2 \tag{8}$$

As shown in Figure 1, liquidity risk is a function of the time τ between clearings for two reasons. First, a larger value of τ leads to more participants in a given metting and, based on equation (6), this reduces the transient deviation of the clearing price from the contemporaneous equilibrium value. The induced reduction in liquidity risk is shown in the first term of equation (8). It is this effect which leads transactors to prefer to trade in infrequently clearing, consolidated, markets (or to spend time waiting for the arrival of compatible trading partners). They bear less risk of getting a transiently unfavorable price in such markets. However, larger values of τ, according to equation (5), increase the exposure of a transactor to the market risk of changes in equilibrium values. This is represented by the second term of equation (8).

From equation (8) it is easy to show that the time between clearings which

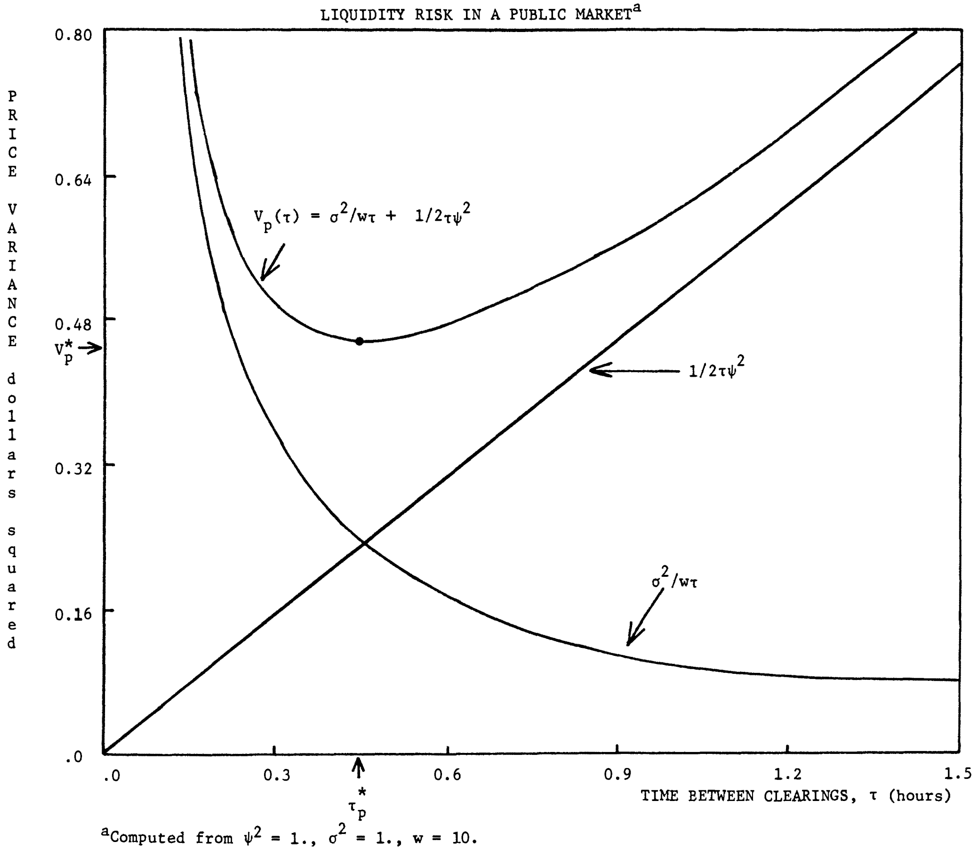


Figure 1. Liquidity Risk in a Public Market[a]

minimizes total liquidity risk is:

$$\tau_p^* = (2/w)^{1/2}\sigma/\psi \tag{9}$$

Equation (9) shows that the optimal time between clearings is a decreasing function of the size of the market, represented by w, and a decreasing function of the equilibrium price volatility of the security, represented by ψ. A shorter time between clearings means more frequent clearings. In the limit, equation (9) indicates that clearings become almost continuous as market size increases and as equilibrium price volatility increases.

These results yield conjectures regarding real world observations once we assume that investors prefer to trade in markets which minimize their exposure to liquidity risk. The dominant exchange structure for a security will reflect the risk-minimizing clearing frequency for that security. In terms of our previous illustrations, the demand during the Civil War for trading in markets which cleared more than twice a day can be explained as a response to the increased volatility of securities prices (larger ψ). The survival of continuous markets after the Civil War can be attributed to the effect of stock tickers and brokers wires in enlarging the geographic scope of those markets (larger w). The same analysis also provides an explanation for the observation that most transactions in actively traded common stock issues are today completed quickly (their high price volatility implies a small optimal time to clear the order), while favorable transactions in thinly traded municipals with stable equilibrium values may take up to several days to complete.

The minimum liquidity risk borne in trading an asset is the value of V_p in equation (8) evaluated at $\tau = \tau_p^*$. That minimum risk is:

$$V_p^* = \sigma\psi(2/w)^{1/2} \tag{10}$$

Recall, from section 1, that an asset is commonly considered liquid if it trades in a large-sized market or if it has a relatively stable equilibrium value. These two dimensions of liquidity are captured in the expression for V_p^*. In particular, V_p^* is a decreasing function of market size, represented by w, and an increasing function of price volatility, represented by ψ. This shows why actively traded volatile stocks and inactively traded short-term debt instruments with stable prices can both be considered liquid assets. Note, however, that the optimal times taken to execute transactions in the two instruments will be quite different, and that the organization of their respective exchange markets will consequently differ.

IV. Dealer Activity in a Public Market

If participation in a market is restricted to the public transactors identified in the model of section 3, the market clearing price r_t will evolve through time as a random walk with additive and serially uncorrelated transient deviations in each clearing (see equations (5) and (6)). It can be shown, Garbade and Lieber, [5] that clearing prices in such a market will exhibit a reversal structure, so that a price increase, $r_t > r_{t-1}$, would be expected to follow a decline, $r_{t-1} < r_{t-2}$, and conversely. Such reversals imply the existence of profit opportunities. We first consider how market participants can exploit these opportunities in their trading and then turn to the effect of their actions on liquidity risk and market organization.

From equation (6) we have that the clearing price in a public market will sometimes be greater and sometimes smaller than the contemporaneous equilibrium value of the asset, but that the former is an unbiased estimator of the latter. There exists, however, an estimator of m_t with a variance smaller than that of r_t. This creates an incentive for transactors to use this estimator and to enter the market as buyers when r_t is smaller than their estimate of the current equilibrium value and as sellers when it is larger. We first describe the estimation algorithm. For expositional convenience we refer to those who trade on the basis of the algorithm as dealers. There is, however, no reason why any public transactor could not function as a dealer in addition to entering his own orders according to the demand schedule of equation (1). After describing the algorithm we develop the characteristics of a market which includes both dealer and public transactors.

Recursive Estimation of Equilibrium Value

As an estimator of m_t, r_t uses only the most current observation on the market clearing price. However, because successive clearing prices have a temporal dependence, one can estimate m_t more efficiently using the entire past sequence of observations on r_t. The resulting estimation algorithm is known as Kalman filtering, Athans, [1]; Sage and Melsa, [14].

Let $\hat{m}_t$ denote the estimate of m_t based on the observations $(r_t, r_{t-1}, r_{t-2}, \cdots)$, i.e., based on observations of the mean reservation prices up to and including time t. It is shown in an appendix that this estimator follows the recursive process:

$$\hat{m}_t = \hat{m}_{t-1} + G[r_t - \hat{m}_{t-1}] \tag{11a}$$

where:

$$G = 2\tau\psi^2/[\tau\psi^2 + (\tau^2\psi^4 + 4\psi^2\sigma^2/w)^{1/2}] \tag{11b}$$

$\hat{m}_t$ is an unbiased estimator of m_t with estimation variance S where:

$$S = [(\tau^2\psi^4 + 4\psi^2\sigma^2/w)^{1/2} - \tau\psi^2]/2 \tag{11c}$$

The Kalman gain parameter G is bounded from above by unity and is an increasing function of τ, the time between market clearings, and w, the time rate of exposure of public transactors to the market. The estimation variance S is strictly less than $\sigma^2/w\tau$ and approaches that value as either w or τ increase. This shows that $\hat{m}_t$ is more efficient than r_t as an estimator of the equilibrium value m_t at time t.

The Effect of Dealer Activity on Market Clearing Prices

We now show the consequences of dealer trading based on the above algorithm for estimating the equilibrium value of a security. To avoid complications to the analysis we assume the mean reservation price r_t of public investors is observable, even when dealer activity forces a divergence between r_t and the market clearing price.[5] We also assume that dealers as well as public transactors perceive the market as perfectly competitive.

Let $\hat{m}_t$ be the estimate of the equilibrium value of the asset at time t from equation (11a) and let p_t be the market clearing price at time t. All trades at time t are executed at the clearing price. The aggregate gross dealer demand for securities is assumed to be:

$$Q_t = c(\hat{m}_t - p_t) \qquad c > 0 \tag{12}$$

This demand schedule may be a summation of the demand schedules of many dealers. For the present analysis we take c, the magnitude of the dealers' response to discrepancies between their estimate of equilibrium and the market clearing price, as given. In fact, c would be determined competitively by the risk-adjusted rate of return on dealer capital. We expect c to be larger for securities with a large number of active public transactors (w) and for securities with low price volatility (ψ), since both conditions would imply smaller risk exposure of dealer capital.[6]

[5] For dealers (or stock exchange specialists) this might be quite reasonable given the flow of public orders seen directly.

[6] The Special Study of the Securities Markets [18] reports that specialists (p. 86–87) and floor traders (p. 205–207, 220–221) are more active in heavily traded issues than in thinly traded issues. This concurs with our conjecture that c is an increasing function of w.

At time t the securities brought by dealers to the market are their positions Q_{t-1} from the preceeding market meeting. Since the market is competitive, no dealer perceives the potential liquidation of his holdings as affecting the clearing price. This allows us to treat the net change in dealer inventories, $Q_t - Q_{t-1}$, in the auction at time t as a liquidation of their existing inventories Q_{t-1} and the acquisition of a new position Q_t.

The market clears at the price where aggregate supply equals aggregate demand. To the earlier supply/demand balance of equation (2) we now have to add the participation of the dealers. If there are K public transactors the market equilibrium condition is:

$$KE + Q_{t-1} + \sum_{i=1}^{K} [E + a(r_i - p_t)] + c(\hat{m}_t - p_t) \tag{13}$$

The market clearing price is then:

$$p_t = \alpha r_t + (1 - \alpha)\hat{m}_t - \gamma Q_{t-1} \tag{14}$$

where:

$$\alpha = aK/(aK + c)$$

$$\gamma = 1/(aK + c)$$

Substituting p_t into equation (12) we have the quantity of securities absorbed by dealers:

$$Q_t = (1 - \alpha)Q_{t-1} + aKc\gamma(\hat{m}_t - r_t) \tag{15}$$

As above we can pass to the continuum case by replacing the discrete number K of public transactors in equations (14) and (15) with $w\tau$.

Several features of equations (14) and (15) are worth noting. If dealers are carrying no net inventories from time $t - 1$, so that $Q_{t-1} = 0$, the market will clear at a price which is a convex combination of r_t (the price it would have cleared at were there no dealers) and $\hat{m}_t$ (the dealers' estimate of m_t). The weights on r_t and $\hat{m}_t$ depend on the relative slopes of the aggregate demand schedules of the public participants, $aK = aw\tau$, and of the dealers, c. The larger the slope c of the dealers demand schedule, the closer will p_t be to $\hat{m}_t$.[7] From the discussion above $\hat{m}_t$ is a more efficient estimator of m_t than r_t. Thus dealers will eliminate more transient price fluctuations the larger the value of c.

If the dealers are carrying over positive inventories into the market meeting at time t, $Q_{t-1} > 0$, the clearing price p_t will be lower than if $Q_{t-1} = 0$, and conversely if they are carrying over short positions. The existence of dealer inventories moderates the contribution of dealers to the stabilization of the clearing price around the equilibrium value. However, Q_{t-1} will be positive only if the market would have otherwise cleared below $\hat{m}_{t-1}$ at time $t - 1$, i.e., only if $r_{t-1} < \hat{m}_{t-1}$. By buying securities at time $t - 1$ the dealers gave a "fairer" price to market

[7] As the slope c of the dealers' demand schedule increases and α decreases the steady-state variance of Q_t will increase, see equation (15). When c goes to infinity Q_t follows a random walk because the innovation $r_t - \hat{m}_t$ in a serially uncorrelated process (Mehra, [12]). It follows that the dealers can "peg" the market to their estimate of the equilibrium value only by accepting a loss of control over their inventory. This result has also been derived by Garman [8] in a quite different model of dealer behavior.

participants in that meeting, and carried over what they perceived as an excess supply of securities into the meeting at time t. Public participants at time t have the opportunity to participate in part of the excess supply from time $t - 1$ through the purchase of dealers inventories.

It should also be noted that the distribution of p_t collapses to a singular distribution at the contemporaneous equilibrium value m_t with increasing w.[8] As was the case with a public market, the clearing price in a dealer market approaches the concurrent equilibrium value of the asset as the size of the market increases.

Liquidity Risk and the Optimal Clearing Frequency of Dealer Markets

We can now compare the liquidity risk and optimal clearing frequency of dealer markets with that of pure public markets. When only public participants transact in the market, the difference between the market clearing price (which then equals r_t) and the contemporaneous equilibrium value m_t is, by equation (6), a serially uncorrelated process with mean zero and variance $\sigma^2/^{w\tau}$. When dealers enter the market the dynamics of the market clearing price are more complex and are developed in detail in an appendix.

As before, liquidity risk is measured by the variance of the difference between equilibrium prices and subsequent transactions prices. For an investor who decides to trade at time $t - ½$, for execution at time t, the variance is:

$$\begin{aligned}\mathrm{Var}[p_t - m_{t-1/2}] &= \mathrm{Var}[(p_t - m_t) + (m_t - m_{t-1/2})]\\ &= \mathrm{Var}[p_t - m_t] + \mathrm{Var}[m_t - m_{t-1/2}]\\ &\quad + 2\,\mathrm{Cov}[p_t - m_t, m_t - m_{t-1/2}] \qquad (16)\end{aligned}$$

The covariance term in equation (16) is zero when only public investors are active in the market, but will be non-zero when dealers are present. The reason is that p_t depends on $\hat{m}_t$, and $\hat{m}_t$ is not statistically independent of the equilibrium price change $m_t - m_{t-1/2}$.

The liquidity risk measure for a dealer market given in equation (16) can be compared to that of equation (7) for the case of a public market. Let $\theta(\tau)$ denote the steady-state value of $\mathrm{Var}[p_t - m_t]$. If dealers are inactive, so that $c = 0$, $\theta(\tau)$ will equal $\sigma^2/w\tau$. For values of c greater than zero $\theta(\tau)$ will be less than $\sigma^2/w\tau$ because of the price stabilizing effect of dealer intervention (See Figure 2). It is shown in an appendix that the covariance term in (16) is equal to ½ $(1 - \alpha)$ $(G - 1)\tau\psi^2$. The total liquidity risk of trading in a dealer market is therefore:

$$V_d = \theta(\tau) + ½\,\tau\psi^2 + (1 - \alpha)(G - 1)\tau\psi^2 \qquad (17)$$

Comparing equations (8) and (17) we have that, for any value of τ, the liquidity risk of trading in a dealer market is less than that of trading in a market open only to public transactors, i.e., $V_d < V_p$. This follows because $\theta(\tau) < \sigma^2/w\tau$ and

[8] This occurs because in equation (14) γ goes to zero, $\hat{m}_t$ goes to r_t and, from equation (6), the distribution of r_t collapses to a singular distribution at m_t with increasing w. Hence the distribution of p_t collapses to a singular distribution at m_t with increasing w. $\hat{m}_t$ converged to r_t with increasing w because the Kalman gain coefficient G converged to unity with increasing w, see equations (11a) and (11b).

because $G < 1$ and $\alpha < 1$ for all values of τ. Figure 2 compares the behavior of the liquidity risk measures V_d and V_d as functions of τ, and also shows the behavior of the determinants of V_d from equation (16). Because $V_d < V_p$ for all τ, we anticipate that public investors will prefer to trade in dealer markets. In particular, if there were two competing markets, alike in all respects except one permits and one prohibits dealers, we expect that the former market would gain public participants at the expense of the latter.

We can show also that dealer participation in a market reduces the optimal time between clearings, or increases the optimal frequency of clearing. Thus, liquidity risk in dealer markets is lower than that for public markets for all values of τ and, in addition, the value of τ which minimizes liquidity risk in a dealer market is smaller than that which minimizes liquidity risk in a purely public market. The latter assertion may be verified by noting that the derivative of V_d with respect to τ, denoted V'_d, is positive at τ_p^* taken from equation (9), the value of τ which minimizes liquidity risk in a public market.[9] This behavior of V_d is noted in Figure 2.

If the slope c of the dealers demand schedule is large, it may be that the minimum value of V_d occurs at $\tau = 0$. In this case the market literally clears continuously, so that all public orders are executed immediately upon their arrival in the market. For smaller values of c market participants will find waiting for the arrival of compatible trading partners in their self-interest.

We have shown thus far that dealers improve the liquidity of markets in two ways. First, for any given time interval between clearings they reduce the liquidity risk borne by public transactions. Second, they also reduce the optimal time between clearings (or increase the optimal frequency of clearing) and in so doing reduce liquidity risk even further. The magnitude of these effects depends upon how much dealers buy when the market price would otherwise clear below their estimate of equilibrium and how much dealers sell when it would otherwise clear above. In other words, the slope c of the dealer demand schedule in equation (12) determines the magnitude of dealer impact on market structure and on the liquidity characteristics of the asset being traded.

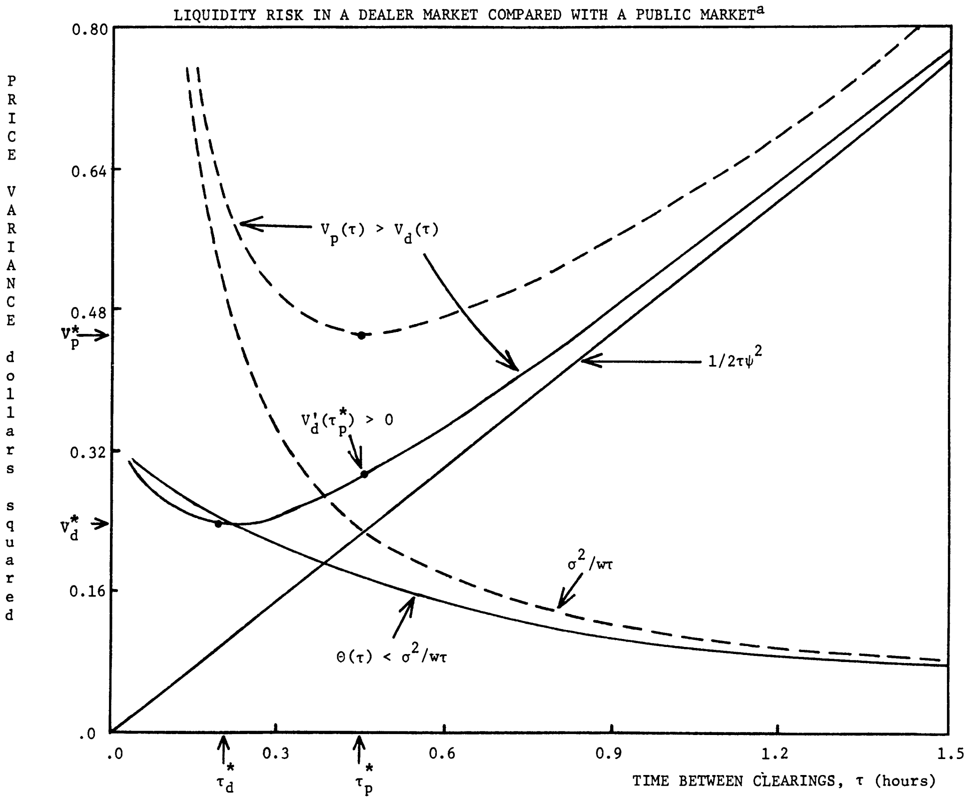
We suggested above that the extent of dealer participation, c, would be determined in a general equilibrium model by the risk-adjusted competitive rate of return on dealer capital. While such a treatment is beyond the scope of this paper, one point may be noted. Technological devices which allow dealers to expand their contacts with potential transactors should increase the level of

[9] From equation (17) we have:

$$V'_d = \theta' + \tfrac{1}{2}\,\psi^2 + (1-\alpha)(G-1)\psi^2 - \alpha'(G-1)\tau\psi^2 + G'(1-\alpha)\tau\psi^2$$

$$= [\theta' + \tfrac{1}{2}\,\psi^2] + [1-\alpha-\tau\alpha'](G-1)\psi^2 + [G'](1-\alpha)\tau\psi^2$$

V'_d will be positive if each of the bracketed terms in the above equation are positive. Since the Kalman gain parameter is a monotonically increasing function of τ we have $G' > 0$. From the definition of $\alpha = aw\tau/(aw\tau + c)$ it follows that $1 - \alpha - \tau\alpha' = c^2/(aw\tau + c)^2 > 0$. As noted in the text, $\theta < \sigma^2/w\tau$ for all $\tau > 0$ because of the stabilizing effect of dealer intervention in the market. As τ increases, however, the importance of public participation relative to dealer participation increases, so that $d(\sigma^2/w\tau)/d\tau < \theta' < 0$ (see Figure 2) or $-\sigma^2/w\tau^2 < \theta'$. At $\tau = \tau_p{}^*$ we have $\tfrac{1}{2}\,\psi^2 = \sigma^2/w\tau^2$ (see equation (9)) so that $\theta' + \tfrac{1}{2}\,\psi^2 > 0$ at $\tau = \tau_p^*$. It follows that $V'_d > 0$ at $\tau = \tau_p^*$.



[a]Computed from ψ^2 = 1., σ^2 = 1., w = 10., a = 10., c = 500.

Figure 2. Liquidity Risk in a Dealer Market Compared with a Public Market[a]

dealer participation. Communications and information storage technologies which reduce the riskiness of dealer positions will increase the value of c and therefore improve further the liquidity characteristics of dealer markets.

The development of on-line data bases disclosing the portfolio holdings of institutional investors and permitting the recall of the purchase and sale interests of those investors is an area of major interest among securities firms active in the corporate bond markets. In addition, the "blue list" of purchase and sale interests in municipal bonds, published by Standard and Poors, is currently being converted from a daily printed listing to a computer listing disclosed through CRT devices with a capability for instantaneous updating. These devices are likely to lead to improved liquidity in these markets.

V. Concluding Remarks

This paper departs from the standard analysis of secondary markets by treating the structural organization of exchange endogenously. We hypothesized that, through competition between markets with different organizational characteristics, a surviving market will be organized to minimize trader exposure to liquidity risk. The optimal frequency of market clearing was shown to depend upon the volatility of securities prices and the number of market participants. Dealer participation in market trading reduces liquidity risk and increases the optimal frequency of clearing. Advances in communications and information technology have similar effects. The empirical relevance of these results was illustrated by a number of historical and contemporary examples of the structure of secondary markets.

Appendix A: Recursive Estimation of Equilibrium Values

The structure of equation (11) is a direct application of Kalman filtering, Athans, [1]; Sage and Melsa, [14]. From equations (5) and (6) we have:

$$m_t = m_{t-1} + e_t \qquad e_t \sim N(0, \tau\psi^2) \tag{A1a}$$

$$r_t = m_t + f_t \qquad f_t \sim N(0, \sigma^2/w\tau) \tag{A1b}$$

The problem is to estimate m_t using the concurrent and past values of r_t.

Let $\hat{m}_t$ be the estimate of m_t based on $(r_t, r_{t-1}, r_{t-2}, \cdots)$ and let $\tilde{m}_t$ be the estimate of m_t based on $(r_{t-1}, r_{t-2}, \cdots)$. Then from (A1) we have:

$$\tilde{m}_t = \hat{m}_{t-1} \tag{A2a}$$

$$\hat{m}_t = \tilde{m}_t + G_t[r_t - \tilde{m}_t] \tag{A2b}$$

where:

$$G_t = R_t/(R_t + \sigma^2/w\tau) \tag{A3a}$$

$$R_t = S_{t-1} + \tau\psi^2 \tag{A3b}$$

$$S_t = R_t - G_t R_t \tag{A3c}$$

It can be shown that the estimators are distributed as:

$$\tilde{m}_t \sim N(m_t, R_t) \tag{A4a}$$

$$\hat{m}_t \sim N(m_t, S_t) \tag{A4b}$$

As t increases G_t, R_t and S_t will converge to steady-state values denoted by G, R and S. These values may be computed by solving the set of simultaneous equations:

$$G = R/(R + \sigma^2/w\tau) \tag{A5a}$$

$$R = S + \tau\psi^2 \tag{A5b}$$

$$S = R - GR \tag{A5c}$$

Thus we have:

$$\begin{aligned} R &= S + \tau\psi^2 \\ &= [R - GR] + \tau\psi^2 \\ &= R - [R/(R + \sigma^2/w\tau)]R + \tau\psi^2 \end{aligned}$$

or:

$$R^2 - \tau\psi^2 R - \psi^2\sigma^2/w = 0$$

Solving the quadratic gives:

$$R = \tfrac{1}{2}[\tau\psi^2 + (\tau^2\psi^4 + 4\psi^2\sigma^2/w)^{1/2}] \tag{A6}$$

since the other root is negative. The value of S follows immediately as $S = R - \tau\psi^2$. The value of G follows from: $GR = R - S = \tau\psi^2$ or $G = \tau\psi^2/R$.

Appendix B: Dynamics of Clearing Prices in a Dealer Market

Four equations summarize the dynamic behavior of clearing prices relative to equilibrium values in a dealer market. From equation (4a) we have:

$$r_t - m_t = f_t \tag{B1}$$

From equation (6a) we have:

$$\begin{aligned}\hat{m}_t - m_t &= (1 - G)\hat{m}_{t-1} + Gr_t - m_{t-1}\\ &= (1 - G)(\hat{m}_{t-1} - m_t) + G(r_t - m_t)\end{aligned}$$

Noting that $m_t = m_{t-1} + e_t$ from equation (5a) this becomes:

$$\hat{m}_t - m_t = (1 - G)(\hat{m}_{t-1} - m_{t-1}) + G(r_t - m_t) - (1 - G)e_t \tag{B2}$$

Equation (14a) may be written in terms of deviations from m_t:

$$p_t - m_t = \alpha(r_t - m_t) + (1 - \alpha)(\hat{m}_t - m_t) - \gamma Q_{t-1} \tag{B3}$$

and equation (15) similarly:

$$Q_t = (1 - \alpha)Q_{t-1} + aKc\gamma(\hat{m}_t - m_t) - aKc\gamma(r_t - m_t) \tag{B4}$$

Equations (B1, B2, B3, B4) provide a system of simultaneous recursive stochastic equations for the price differences $r_t - m_t$, $\hat{m}_t - m_t$ and $p_t - m_t$, and for the quantity Q_t of securities held by positioning agents. These equations describe the transient characteristics of the market. The equations may be summarized in a multi-variate linear dynamic model:

$$A_0 x_t = A_1 x_{t-1} + B_1 \epsilon_t \tag{B5a}$$

$$\epsilon_t \sim N(0, \textstyle\sum) \tag{B5b}$$

with:

$$x_t = \begin{bmatrix} r_t - m_t \\ \hat{m}_t - m_t \\ p_t - m_t \\ Q_t \end{bmatrix} A_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\alpha & \alpha - 1 & 1 & 0 \\ aKc\gamma & -aKc\gamma & 0 & 1 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 - G & 0 & 0 \\ 0 & 0 & 0 & -\gamma \\ 0 & 0 & 0 & 1 - \alpha \end{bmatrix} B_1 = \begin{bmatrix} 1 & 0 \\ G & G - 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\epsilon_t = \begin{bmatrix} f_t \\ e_t \end{bmatrix} \textstyle\sum = \begin{bmatrix} \sigma^2/w\tau & 0 \\ 0 & \tau\psi^2 \end{bmatrix}$$

Note that the random process ϵ_t is serially uncorrelated.

The reduced-form version of (B5a) is:

$$x_t = A_0^{-1}A_1x_{t-1} + A_0^{-1}B_1\epsilon_t \tag{B6}$$

Stationary stochastic models of this type have been studied by Chow, [2], who shows that the steady-state covariance matrix of x_t is the matrix U which satisfies the equation:

$$U = A_0^{-1}A_1U(A_0^{-1}A_1)' + A_0^{-1}B_1 \sum (A_0^{-1}B_1)' \tag{B7}$$

The steady-state value of the variance of $p_t - m_t$, which is denoted $\theta(\tau)$ in the text, is equal to $U_{3,3}$. The steady-state value of the covariance of $p_t - m_t$ with $m_t - m_{t-1/2}$ is $\frac{1}{2}[A_0^{-1}B_1\sum]_{3,2} = [A_0^{-1}B_1]_{3,2}\ (\frac{1}{2}\tau\psi^2)$. It may be verified directly that:

$$A_0^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \alpha & 1-\alpha & 1 & 0 \\ -aKc\gamma & aKc\gamma & 0 & 1 \end{bmatrix}$$

so $[A_0^{-1}B_1]_{3,2} = (1 - \alpha)(G - 1)$.

References

1. Michael Athans. "The Importance of Kalman Filtering Methods for Economic Systems." *Annals of Economic and Social Measurement,* 3 (January, 1974).
2. Gregory Chow. "The Acceleration Principle and the Nature of Business Cycles." *Quarterly Journal of Economics,* 82 (August, 1968).
3. David Eiteman and Wilfod Eiteman. *Leading World Stock Exchanges: Trading Practices and Organization.* Michigan International Business Studies, No. 2, (1964).
4. Kenneth Garbade. "The Effect of Inter-Dealer Brokerage on the Transactional Characteristics of a Dealer Market." *Journal of Business,* 51 (July, 1978).
5. ——— and Zvi Lieber. "On the Independence of Transactions on the New York Stock Exchange." *Journal of Banking and Finance,* 1 (September, 1977).
6. ——— and William Silber. "Price Dispersion in the Government Securities Market." *Journal of Political Economy,* 84 (August, 1976).
7. ———. "Technology, Communication and the Performance of Financial Markets, 1840–1975." *Journal of Finance* , 33 (June, 1978).
8. M. B. Garman. "Market Microstructure." *Journal of Financial Economics,* 3 (June, 1976).
9. Horace Hotchkiss. "The Stock Ticker." in *The New York Stock Exchange,* Edmund Stedman, ed. (New York: Greenwood Press Publishers, 1905).
10. Henry G. Jarecki. "Bullion Dealing, Commodity Exchange Trading and the London Gold Fixing: Three Forms of Commodity Auctions." in *Bidding and Auctioning for Procurement and Allocation,* Yakov Amihud, ed. (New York: New York University Press, 1976).
11. J. E. Meeker. *The Work of the Stock Exchange,* rev. ed. (New York: The Ronald Press Company, 1930).
12. Raman Mehra. "On the Identification of Variances and Adaptive Kalman Filtering." *IEEE Transactions on Automatic Control,* 15 (April, 1970).
13. New York Stock Exchange. *Report of the Committee to Study the Stock Allocation System* (New York: New York Stock Exchange, 1976).
14. Andrew Sage and James Melsa. *Estimation Theory with Applications to Communications and Control* (New York: McGraw-Hill, 1971).
15. William L. Silber. "Thinness in Capital Markets: The Case of the Tel Aviv Stock Exchange." *Journal of Financial and Quantitative Analysis,* 10 (March, 1975).
16. Robert Sobel. *The Big Board, A History of the New York Stock Exchange* (New York: The Free Press, 1965).
17. ———. *N.Y.S.E.* (New York: Weybright and Talley, 1975).
18. U.S. Securities and Exchange Commission. *Report of Special Study of Securities Markets,* Part 2, House Document No. 95, pt. 2, 88th Congress, 1st Session (Washington, D.C.: U.S.G.P.O., 1963).