Econometrica, Vol. 73, No. 6 (November, 2005), 1815–1847

OVER-THE-COUNTER MARKETS

BY DARRELL DUFFIE, NICOLAE GÂRLEANU, AND LASSE HEJE PEDERSEN[1]

We study how intermediation and asset prices in over-the-counter markets are affected by illiquidity associated with search and bargaining. We compute explicitly the prices at which investors trade with each other, as well as marketmakers' bid and ask prices, in a dynamic model with strategic agents. Bid–ask spreads are lower if investors can more easily find other investors or have easier access to multiple marketmakers. With a monopolistic marketmaker, bid–ask spreads are higher if investors have easier access to the marketmaker. We characterize endogenous search and welfare, and discuss empirical implications.

KEYWORDS: Asset pricing, search frictions, bargaining, marketmaking, welfare, Walrasian.

IN OVER-THE-COUNTER MARKETS, an investor who wishes to sell must search for a buyer, incurring opportunity or other costs until one is found. Some over-the-counter (OTC) markets therefore have intermediaries. Contact with relevant intermediaries, however, is not immediate. Often, intermediaries must be approached sequentially. Hence, when two counterparties meet, their bilateral relationship is inherently strategic. Prices are set through a bargaining process that reflects each investor's or marketmaker's alternatives to immediate trade.

These search-and-bargaining features are empirically relevant in many markets, such as those for mortgage-backed securities, corporate bonds, emerging-market debt, bank loans, derivatives, and certain equity markets. In real-estate markets, for example, prices are influenced by imperfect search, the relative impatience of investors for liquidity, outside options for trade, and the role and profitability of brokers.

We build a dynamic asset-pricing model that captures these features and analytically derive the equilibrium allocations, prices negotiated between investors, as well as marketmakers' bid and ask prices. We show how these equilibrium properties depend on investors' search abilities, marketmaker accessibility, and bargaining powers. We determine the search intensities that marketmakers choose, and derive the associated welfare implications of investment in marketmaking.

Our model of search is a variant of the coconuts model of Diamond (1982).[2] A continuum of investors contact each other, independently, at some mean

[1] Part of this paper was previously distributed under the title "Valuation in Dynamic Bargaining Markets." We are grateful for conversations with Yakov Amihud, Helmut Bester, Joseph Langsam of Morgan Stanley Dean Witter, Richard Lyons, Tano Santos, and Jeff Zwiebel, and to participants at the NBER Asset Pricing Meeting, the Cowles Foundation Incomplete Markets and Strategic Games Conference, the Western Finance Association Conference, the CEPR meeting at Gerzensee, University College London, Universite Libre de Bruxelles, Tel Aviv University, and Universitat Autonoma de Barcelona.

[2] Our model differs from Diamond (1982), and the labor literature more generally, by considering repeated trade of long-lived assets. The monetary search literature (for example,

rate λ, a parameter that reflects search ability. Similarly, marketmakers contact agents at some intensity ρ that reflects dealer availability. When agents meet, they bargain over the terms of trade. Gains from trade arise from heterogeneous costs or benefits of holding assets. For example, an asset owner can be anxious to sell because of a liquidity need or because of hedging motives. Marketmakers are assumed to off-load their inventories in a frictionless interdealer market, trading with investors so as to capture part of the difference between the interdealer price and investors' reservation values.

Asset pricing with exogenous trading frictions has been studied by Amihud and Mendelson (1986), Constantinides (1986), and Vayanos (1998). We endogenize the trading frictions that arise through search and bargaining, and show their effects on asset prices. In follow-up work, Duffie, Gârleanu, and Pedersen (2003) extend the model developed here to characterize the impact on asset pricing of search in settings with risk aversion and risk limits, while Weill (2002) and Vayanos and Wang (2002) consider cross-sectional asset pricing in extensions with multiple assets.

Market frictions have been used to explain the existence and behavior of marketmakers. Notably, marketmakers' bid and ask prices have been explained by inventory considerations (Garman (1976), Amihud and Mendelson (1980), and Ho and Stoll (1981)) and by adverse selection arising from asymmetric information (Bagehot (1971), Glosten and Milgrom (1985), and Kyle (1985)). In contrast, we model marketmakers who have no inventory risk because of the existence of interdealer markets, and our agents are symmetrically informed. In our model, bid and ask prices are set in light of investors' outside options, which reflect both the accessibility of other marketmakers and investors' own abilities to find counterparties.

We show that bid–ask spreads are lower if investors can find each other more easily.[3] The intuition is that improving an investor's search alternatives forces marketmakers to give better prices. This result is supported by the experimental evidence of Lamoureux and Schnitzlein (1997).

An investor also improves his bargaining position relative to a marketmaker if he can more easily find other marketmakers. Hence, despite the bilateral nature of bargaining between a marketmaker and an investor, marketmakers are effectively in competition with each other over order flow, given the option

Kiyotaki and Wright (1993)) also considers long-lived assets, but, with the exception of Trejos and Wright (1995), it considers exogenous prices. Our model has similarities with that of Trejos and Wright (1995), but their objectives are different and they do not study marketmaking. See also Harris (1979).

[3] We show that our model specializes in a specific way to the standard general-equilibrium paradigm as bilateral trade becomes increasingly active (under conditions to be described), extending a chain of results by Rubinstein and Wolinsky (1985), Gale (1987, 1986a, 1986b), and McLennan and Sonnenschein (1991), in a manner explained later in our paper. Thus, "standard" asset-pricing theory is not excluded, but rather is found at the end of the spectrum of increasingly "active" markets.

of investors to search for better terms. Consistent with this intuition, we prove that competitive prices and vanishing spreads obtain as marketmakers' contact intensities become large, provided that marketmakers do not have all of the bargaining power.

In summary, if investors are more sophisticated (that is, have better access to other investors or to marketmakers who do not have total bargaining power), they receive a *tighter* bid–ask spread. This implication sets our theory of intermediation apart from information-based models, in which more sophisticated (that is, better informed) investors receive a wider bid–ask spread.

In an extension with heterogeneous investors in the same OTC market, we show that more sophisticated investors (those with better access to marketmakers) receive tighter bid–ask spreads because of their improved outside options. Hence, this result holds both when comparing across markets and when comparing across investors in the same market. This sets our theory apart from inventory-based models, which would not imply differential treatment across investors.[4] Furthermore, in the heterogeneous-agents extension, investors with lower search ability may refrain entirely from trade.

Our result seems consistent with behavior in certain OTC markets, such as those for interest-rate swaps and foreign exchange, in which asymmetric information is limited. Anecdotal evidence suggests that "sales traders" give more competitive prices to sophisticated investors, perceived to have better outside options.

We also consider cases in which the marketmaker has total bargaining power. The bid–ask spread of such a monopolistic marketmaker vanishes as investors are increasingly able to meet each other quickly, as with the case of competing marketmakers. In contrast, however, more frequent contact between investors and a monopolistic marketmaker actually widens spreads, because of the investors' poorer outside options. Specifically, an investor's threat to find a counterparty himself is less credible if the marketmaker has already executed most of the efficient trades, making it harder for the investor to find potential counterparties.

Our results regarding the impact of investors' searches for each other on dealer spreads are similar in spirit to those of Gehrig (1993) and Yavaş (1996), who consider monopolistic marketmaking in one-period models.[5] We show that dynamics have an important effect on agents' bargaining positions, and thus asset prices, bid–ask spreads, and investments in marketmaking capacity. Rubinstein and Wolinsky (1987) study the complementary effects of marketmaker inventory and consignment agreements in a dynamic search model.

[4] We note that, when comparing across markets, inventory considerations may have the same bid–ask implication as our search model, because more frequent meetings between investors and marketmakers may result in lower inventory costs.

[5] See also Bhattacharya and Hagerty (1987), who introduce dealers into the Diamond (1982) model, and Moresi (1991), who considers intermediation in a search model in which buyers and sellers exit the market after they trade.

We consider marketmakers' choices of search intensity and the social efficiency of these choices. A monopolistic marketmaker imposes additional "networking losses" on investors because his intermediation renders less valuable the opportunity of investors to trade directly with each other. A monopolistic marketmaker thus provides more intermediation than is socially efficient. Competitive marketmakers may provide even more intermediation, because they do not consider, in their allocation of resources to search, the impact of their intermediation on the equilibrium allocation of assets among investors.[6]

1. MODEL

We fix a probability space $(\Omega, \mathcal{F}, \Pr)$ and a filtration $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-algebras satisfying the usual conditions, as defined by Protter (1990). The filtration represents the resolution over time of information commonly available to agents.

Two kinds of agents, investors and marketmakers, consume a single nonstorable consumption good that is used as a numeraire. All agents are risk-neutral and infinitely lived, with time preferences determined by a constant discount rate $r > 0$. Marketmakers hold no inventory and maximize profits.

Investors have access to a risk-free bank account with interest rate r and to an OTC market for a "consol," meaning an asset paying dividends at the constant rate of 1 unit of consumption per year. (Duffie, Gârleanu, and Pedersen (2003) consider extensions with risky securities and risk-averse investors.) The consol can be traded only when an investor finds another investor or a marketmaker, according to a random search model described below. The bank account can also be viewed as a liquid security that can be traded instantly. We require that the value W_t of an investor's bank account be bounded below, ruling out Ponzi schemes.

A fraction s of investors are initially endowed with 1 unit of the asset. Investors can hold at most 1 unit of the asset and cannot short-sell. Because agents have linear utility, we can restrict attention to equilibria in which, at any given time and state of the world, an investor holds either 0 or 1 unit of the asset.

An investor is characterized by whether he owns the asset or not, and by an intrinsic type that is "high" or "low." A low-type investor, when owning the asset, has a holding cost of δ per time unit; a high-type investor has no such holding cost. There are multiple interpretations of the investor types. For instance, a low-type investor may have (i) low liquidity (that is, a need for cash),

[6] Studying endogenous search in labor markets, Mortensen (1982) and Hosios (1990) find that agents may choose inefficient search levels because they do not internalize the gains from trade realized by future trading partners. Moen (1997) shows that search markets can be efficient under certain conditions.

(ii) high financing costs, (iii) hedging reasons to sell,[7] (iv) a relative tax disadvantage,[8] or (v) a lower personal use of the asset. Any investor's intrinsic type switches from low to high with intensity λ_u and switches back with intensity λ_d. For any pair of investors, their intrinsic-type processes are assumed to be independent.

The full set of investor types is $\mathcal{T} = \{ho, hn, lo, ln\}$, where the letters "$h$" and "$l$" designate the investor's intrinsic liquidity state, as above, and "o" and "n" indicate whether the investor owns the asset or not, respectively.

We suppose that there is a "continuum" (a nonatomic finite-measure space) of investors and we let $\mu_\sigma(t)$ denote the fraction at time t of investors of type $\sigma \in \mathcal{T}$. Because the fractions of each type of investor add to 1 at any time t,

$$(1) \qquad \mu_{ho}(t) + \mu_{hn}(t) + \mu_{lo}(t) + \mu_{ln}(t) = 1.$$

Because the total fraction of investors owning an asset is s,

$$(2) \qquad \mu_{ho}(t) + \mu_{lo}(t) = s.$$

A pair of investors can negotiate a trade of the consol whenever they meet, for a mutually agreeable number of units of current consumption. (The determination of the terms of trade is to be addressed later.) Investors meet, however, only at random times, in a manner idealized as independent random search, as follows. At the successive event times of a Poisson process with some intensity parameter λ, an investor contacts another agent, chosen from the entire population "at random," meaning with a uniform distribution across the investor population. An investor therefore contacts an investor from a given set D of investors that contains a fraction μ_D of the total population with the mean intensity $\lambda\mu_D$. The total rate at which a group C of independently searching investors of mass μ_C contacts group D investors is almost surely $\mu_C\lambda\mu_D$. Because group D investors contact C investors at the same total rate, the total meeting rate between the two groups is almost surely $2\lambda\mu_C\mu_D$. This assumes that searches are independent in a sense appropriate for an application of the exact law of large numbers for random search and matching among a continuum of agents; Duffie and Sun (2004) provide an exact discrete-time theorem and proof.[9] Random switches in intrinsic types are assumed to be independent of the agent matching processes.

[7] Duffie, Gârleanu, and Pedersen (2003) explore this interpretation in an extension with risk aversion.

[8] Dai and Rydqvist (2003) provide a tax example with potential search effects.

[9] The assumed almost sure meeting rate of $2\lambda\mu_C\mu_D$ is the limit meeting rate of an associated discrete-time finite-agent random search model. Ferland and Giroux (2002) prove a more general version of this assertion rigorously. Here is a sketch of the proof in our setting. Suppose that market (n, Δ) has n agents, for whom, given any pair (i, j) of distinct agents, agent i contacts agent j over a discrete-time period of length Δ with probability $p(n, \Delta) = 1 - e^{-\Delta\lambda/n}$ (the probability of an arrival of a Poisson process with intensity λ/n). Suppose that the indicator $\mathbb{1}_{i,j}$ of

There is a unit mass of independent nonatomic marketmakers. Marketmakers are also found through search, implying that an investor must bargain with marketmakers sequentially as they are found. An investor meets a marketmaker with a fixed intensity, ρ, which can be interpreted as the sum of the intensity of investors' search for marketmakers and marketmakers' search for investors.[10] When an investor meets a marketmaker, they bargain over the terms of trade as described in the next section. Marketmakers have access to an immediately accessible interdealer market on which they unload their positions, so that they have no inventory at any time.

The OTC markets without marketmakers are treated by the special case of our model with $\rho = 0$.

2. DYNAMIC SEARCH EQUILIBRIUM WITH COMPETING MARKETMAKERS

In this section, we explicitly compute the allocations and prices that form a dynamic search-and-bargaining equilibrium. In particular, we compute prices negotiated directly between investors, marketmakers' bid and ask prices, and the interdealer price.

In equilibrium, low-type asset owners want to sell and high-type nonowners want to buy. When two such agents meet, they bargain over the price. Similarly, when investors meet a marketmaker, they bargain over the price. An investor's bargaining position depends on his outside option, which in turn depends on the availability of other counterparties, both now and in the future, and a marketmaker's bargaining position depends on the interdealer price. In deriving the equilibrium, we rely on the insight from bargaining theory that trade hap-

successful contact of j by i is independent across all distinct pairs (i, j) of distinct agents. The mean rate of contact per unit of time of a specific investor with other investors in the (n, Δ) market is $E(\Delta^{-1}\sum_{j\neq i}\mathbb{1}_{i,j}) = \Delta^{-1}(n-1)p(n, \Delta)$, which converges to λ, as in our continuous-time model, as $(n, \Delta) \to (+\infty, 0)$. The per capita total rate of contact per unit of time by a subset $C(n) \subset \{1, \dots, n\}$ that contains a fraction μ_C of the total population with a disjoint subset $D(n)$ that contains a fraction μ_D of the population is

$$S(n, \Delta) = \frac{1}{n\Delta}\left(\sum_{i\in C(n), j\in D(n)} \mathbb{1}_{i,j} + \sum_{i\in D(n), j\in C(n)} \mathbb{1}_{i,j}\right),$$

which has mean $(n\Delta)^{-1}2p(n, \Delta)|C(n)| \cdot |D(n)|$, which converges to $2\lambda\mu_C\mu_D$ as $(n, \Delta) \to (+\infty, 0)$. By the weak law of large numbers (Theorem 6.2 of Billingsley (1986)), $S(n, \Delta)$ converges in probability as $(n, \Delta) \to (+\infty, 0)$ to its expectation, $2\lambda\mu_C\mu_D$, given that $S(n, \Delta)$ is the sum of a divergent number of independent variables whose total variance is shrinking to zero. One caveat is that, in a discrete-time model, an agent can contact more than one other agent at the same time. In that case, an elimination rule can be used to keep only one-to-one matches, but since the probability of contacting more than one agent during a period of length Δ is of the order Δ^2, the meeting rate is as derived above. (The same result holds in the limit even if $C(n)$ and $D(n)$ are not disjoint, but one must make slight (order $1/n$) adjustments to the mean of $S(n, \Delta)$ for overlap in the two groups.)

[10] It would be equivalent to have a mass k of dealers with contact intensity ρ/k, for any $k > 0$.

pens instantly.[11] This allows us to derive a dynamic equilibrium in two steps. First, we derive the equilibrium masses of the different investor types. Second, we compute agents' value functions and transaction prices (taking as given the masses of the investor types).

Assuming, as discussed in the previous section, that the law of large numbers applies, the rate of change of the mass $\mu_{lo}(t)$ of low-type owners is almost surely

$$\dot{\mu}_{lo}(t) = -\big(2\lambda\mu_{hn}(t)\mu_{lo}(t) + \rho\mu_m(t)\big) - \lambda_u\mu_{lo}(t) + \lambda_d\mu_{ho}(t), \tag{3}$$

where $\mu_m(t) = \min\{\mu_{lo}(t), \mu_{hn}(t)\}$. The first term in (3) reflects the fact that agents of type *hn* contact those of type *lo* at a total rate of $\lambda\mu_{hn}(t)\mu_{lo}(t)$, while agents of type *lo* contact those of type *hn* at the same total rate $\lambda\mu_{hn}(t)\mu_{lo}(t)$. At both of these types of encounters, the agent of type *lo* becomes one of type *ln*. This implies a total rate of reduction of mass due to these encounters of $2\lambda\mu_{hn}(t)\mu_{lo}(t)$. Similarly, investors of type *lo* meet marketmakers with a total contact intensity of $\rho\mu_{lo}(t)$. If $\mu_{lo}(t) \leq \mu_{hn}(t)$, then all of these meetings lead to trade and the *lo* agent becomes an *ln* agent, resulting in a reduction in μ_{lo} of $\rho\mu_{lo}(t)$. If $\mu_{lo}(t) > \mu_{hn}(t)$, then not all these meetings result in trade. This is because marketmakers buy from *lo* investors and sell to *hn* investors, and, in equilibrium, the total intensity of selling must equal the intensity of buying. Marketmakers meet *lo* investors with total intensity $\rho\mu_{lo}$ and *hn* investors with total intensity $\rho\mu_{hn}$, and, therefore, investors on the "long side" of the market are rationed. In particular, if $\mu_{lo}(t) > \mu_{hn}(t)$, then *lo* agents trade with marketmakers only at the intensity $\rho\mu_{hn}$. In equilibrium, this rationing can be the outcome of bargaining because the marketmaker's reservation value (that is, the interdealer price) is equal to the reservation value of the *lo* investor.

Finally, the term $\lambda_u\mu_{lo}(t)$ in (3) reflects the migration of owners from low to high intrinsic types, and the last term $\lambda_d\mu_{ho}(t)$ reflects owners' change from high to low intrinsic types.

The rate of change of the other investor-type masses are

$$\dot{\mu}_{hn}(t) = -\big(2\lambda\mu_{hn}(t)\mu_{lo}(t) + \rho\mu_m(t)\big) + \lambda_u\mu_{ln}(t) - \lambda_d\mu_{hn}(t), \tag{4}$$

$$\dot{\mu}_{ho}(t) = \big(2\lambda\mu_{hn}(t)\mu_{lo}(t) + \rho\mu_m(t)\big) + \lambda_u\mu_{lo}(t) - \lambda_d\mu_{ho}(t), \tag{5}$$

$$\dot{\mu}_{ln}(t) = \big(2\lambda\mu_{hn}(t)\mu_{lo}(t) + \rho\mu_m(t)\big) - \lambda_u\mu_{ln}(t) + \lambda_d\mu_{hn}(t). \tag{6}$$

As in (3), the first terms reflect the result of trade and the last two terms are the result of intrinsic-type changes.

[11] In general, bargaining leads to instant trade when agents do not have asymmetric information. Otherwise there can be strategic delay. In our model, it does not matter whether agents have private information about their own type, for it is common knowledge that a gain from trade arises only between agents of types *lo* and *hn*.

In most of the paper we focus on stationary equilibria, that is, equilibria in which the masses are constant. In our welfare analysis, however, it is more natural to take the initial masses as given and, therefore, we develop some results with any initial mass distribution. The following proposition asserts the existence, uniqueness, and stability of the steady state.

PROPOSITION 1: *There exists a unique constant steady-state solution to* (1)–(6). *From any initial condition* $\mu(0) \in [0,1]^4$ *that satisfies* (1) *and* (2), *the unique solution* $\mu(t)$ *to* (3)–(6) *converges to the steady state as* $t \to \infty$.

A particular agent's type process $\{\sigma(t) : -\infty < t < +\infty\}$ is, in steady state, a four-state Markov chain with state space $\mathcal{T}$ and with constant switching intensities determined in the obvious way[12] by the steady-state population masses μ and the intensities λ, λ_u, and λ_d. The unique stationary distribution of any agent's type process coincides with the cross-sectional distribution μ of types characterized[13] in Proposition 1.

With these equilibrium masses, we will determine the price P_t negotiated directly between *lo* and *hn* investors, the "bid" price B_t at which investors sell to marketmakers, the "ask" price A_t at which investors buy from marketmakers, and the interdealer price. For this, we use dynamic programming, by first computing an investor's utility at time t for remaining lifetime consumption. For a particular agent this "value function" depends, naturally, only on the agent's current type $\sigma(t) \in \mathcal{T}$, the current wealth W_t in his bank account, and time. More specifically, the value function is

$$U(W_t, \sigma(t), t) = \sup_{C,\theta} E_t \int_0^\infty e^{-rs}\, dC_{t+s} \tag{7}$$

$$\text{subject to} \quad dW_t = rW_t\, dt - dC_t + \theta_t(1 - \delta \mathbb{1}_{\{\sigma^\theta(t) = lo\}})\, dt - \hat{P}_t\, d\theta_t, \tag{8}$$

where E_t denotes $\mathcal{F}_t$-conditional expectation, C is a cumulative consumption process, $\theta_t \in \{0,1\}$ is a feasible asset holding process, σ^θ is the type process induced by θ, and at the time t of a trading opportunity, $\hat{P}_t \in \{P_t, A_t, B_t\}$ is the

[12] For example, the transition intensity from state *lo* to state *ho* is λ_u, the transition intensity from state *lo* to state *ln* is $2\lambda\mu_{hn}$, and so on, for the 4×3 switching intensities.

[13] This is a result of the law of large numbers, in the form of Theorem C of Sun (2000), which provides the construction of our probability space $(\Omega, \mathcal{F}, \Pr)$ and agent space $[0,1]$, with an appropriate σ-algebra making $\Omega \times [0,1]$ into what Sun calls a "rich space," with the properties that: (i) for each individual agent in $[0,1]$, the agent's type process is indeed a Markov chain in $\mathcal{T}$ with the specified generator, (ii) the unconditional probability distribution of the agents' type is always the steady-state distribution μ on $\mathcal{T}$ given by Proposition 1, (iii) agents' type transitions are almost everywhere pairwise independent, and (iv) the cross-sectional distribution of types is also given by μ, almost surely, at each time t.

trade price, which depends on the type of counterparty. From (7) and (8) the value function is linear in wealth, in that $U(W_t, \sigma(t), t) = W_t + V_{\sigma(t)}(t)$, where[14]

$$V_{\sigma(t)}(t) = \sup_{\theta} E_t\left[\int_t^{\infty} e^{-r(s-t)}\theta_s(1 - \delta \mathbb{1}_{\{\sigma^{\theta}(s)=lo\}})\, ds - e^{-r(s-t)}\hat{P}_s\, d\theta_s\right]. \tag{9}$$

As shown in the Appendix, the value functions satisfy the Hamilton–Jacobi–Bellman (HJB) equations

$$\begin{aligned} \dot{V}_{lo} &= rV_{lo} - \lambda_u(V_{ho} - V_{lo}) - 2\lambda\mu_{hn}(P + V_{ln} - V_{lo}) \\ &\quad - \rho(B + V_{ln} - V_{lo}) - (1 - \delta), \\ \dot{V}_{ln} &= rV_{ln} - \lambda_u(V_{hn} - V_{ln}), \\ \dot{V}_{ho} &= rV_{ho} - \lambda_d(V_{lo} - V_{ho}) - 1, \\ \dot{V}_{hn} &= rV_{hn} - \lambda_d(V_{ln} - V_{hn}) - 2\lambda\mu_{ho}(V_{ho} - V_{hn} - P) \\ &\quad - \rho(V_{ho} - V_{hn} - A), \end{aligned} \tag{10}$$

suppressing the time argument t, which implies that an *lo* investor benefits from a sale at any price greater than $V_{lo} - V_{ln}$ and that an *hn* investor benefits from a purchase at any price smaller than $V_{ho} - V_{hn}$. Bargaining between the investors leads to a price between these two values. Specifically, Nash (1950) bargaining with a seller bargaining power of $q \in [0, 1]$ yields

$$P = (V_{lo} - V_{ln})(1 - q) + (V_{ho} - V_{hn})q. \tag{11}$$

This is also the outcome of the simultaneous-offer bargaining game described in Kreps (1990) and of the alternating-offer bargaining game described in Duffie, Gârleanu, and Pedersen (2003).[15]

Similarly, the bid and ask prices are determined through a bargaining encounter between investors and marketmakers in which a marketmaker's outside option is to trade in the interdealer market at a price of M. Marketmakers have a fraction, $z \in [0, 1]$, of the bargaining power when facing an investor. Hence, a marketmaker buys from an investor at the bid price B, and sells at the ask price A, determined by

$$A = (V_{ho} - V_{hn})z + M(1 - z), \tag{12}$$

$$B = (V_{lo} - V_{ln})z + M(1 - z). \tag{13}$$

[14] If $\lim_{s\to\infty} E_t[e^{-rs}\max\{P_s, A_s, B_s\}] = 0$, V is well defined. We restrict attention to such prices.

[15] Duffie, Gârleanu, and Pedersen (2003) describe an alternating-offer bargaining procedure that yields a bargaining power that, in the limit as the time between offers approaches zero, is equal to the probability of making an offer. Our qualitative results do not, however, depend on zero time between offers. For example, the results in Section 4 concerning $\lambda \to \infty$ hold for an arbitrary delay between offers.

As discussed above, in equilibrium, marketmakers and those investors on the long side of the market must be indifferent to trading. Hence, if $\mu_{lo} < \mu_{hn}$, marketmakers meet more potential buyers than sellers. The interdealer price, M, is therefore equal to the ask price A and equal to any buyer's reservation value $V_{ho} - V_{hn}$. Similarly, if $\mu_{lo} > \mu_{hn}$, then $M = B = V_{lo} - V_{ln}$. For the knife-edge case of $\mu_{lo} = \mu_{hn}$, let $M = \tilde{q}(V_{ho} - V_{hn}) + (1 - \tilde{q})(V_{lo} - V_{ln})$, for some constant $\tilde{q}$ that is arbitrarily chosen from [0, 1], and fixed for the remainder.

In steady state, it is easy to see which side of the market is rationed because the steady-state fraction of high-type agents is $\lambda_u(\lambda_d + \lambda_u)^{-1}$, so we have

$$\mu_{hn} + (s - \mu_{lo}) = \frac{\lambda_u}{\lambda_d + \lambda_u}.$$

Hence, $\mu_{lo} < \mu_{hn}$ in steady state if and only if the following condition is satisfied.

CONDITION 1: It holds that $s < \lambda_u/(\lambda_u + \lambda_d)$.

An equilibrium is defined as a process (P, A, B, μ, V) such that (i) the system μ of investor masses solves (1)–(6), (ii) the transaction prices (P, A, B) are those in (11)–(13), and (iii) V solves the HJB equations (9) and (10) and $V_{lo} - V_{ln} \leq V_{ho} - V_{hn}$. As there is a continuum of agents, no agent has the ability to influence mass dynamics with an off-equilibrium-path trading strategy. These three conditions therefore ensure individual-agent optimality of the associated equilibrium trading strategies, as well as consistency between the mass dynamics assumed by agents and those arising from the equilibrium trading strategies. We derive the equilibrium explicitly. For brevity, we report only the prices under Condition 1; the complementary case is treated in the Appendix.

THEOREM 2: *For any given initial mass distribution* $\mu(0)$, *there exists an equilibrium. There is a unique steady-state equilibrium. Under Condition* 1, *the ask, bid, and interinvestor prices are*

$$A = \frac{1}{r} - \frac{\delta}{r}\frac{\lambda_d + 2\lambda\mu_{lo}(1-q)}{r + \lambda_d + 2\lambda\mu_{lo}(1-q) + \lambda_u + 2\lambda\mu_{hn}q + \rho(1-z)}, \tag{14}$$

$$B = \frac{1}{r} - \frac{\delta}{r}\frac{zr + \lambda_d + 2\lambda\mu_{lo}(1-q)}{r + \lambda_d + 2\lambda\mu_{lo}(1-q) + \lambda_u + 2\lambda\mu_{hn}q + \rho(1-z)}, \tag{15}$$

$$P = \frac{1}{r} - \frac{\delta}{r}\frac{(1-q)r + \lambda_d + 2\lambda\mu_{lo}(1-q)}{r + \lambda_d + 2\lambda\mu_{lo}(1-q) + \lambda_u + 2\lambda\mu_{hn}q + \rho(1-z)}. \tag{16}$$

These explicit prices are intuitive. Each price is the present value, $1/r$, of dividends, reduced by an illiquidity discount. All of these prices decrease in the bargaining power z of the marketmaker, because a higher z makes trading more costly for investors. The prices increase, however, in the ease of meeting a marketmaker (ρ) and in the ease of finding another investor (λ), provided that ρ and λ are large enough. The effect of increasing search intensities is discussed in Section 4.

From Theorem 2, the bid–ask spread $A - B$ is increasing in the marketmaker's bargaining power z. The bid–ask spread is decreasing in λ, since a high λ means that an investor can easily find a counterparty himself, which improves his bargaining position. The bid–ask spread is also decreasing in ρ, provided $z < 1$ and ρ is sufficiently large. A higher ρ implies that an investor can quickly find another marketmaker, and this "sequential competition" improves his bargaining position. If $z = 1$, however, then the bid–ask spread is increasing in ρ. The case of $z = 1$ is perhaps best interpreted as that of a monopolistic marketmaker, as discussed in the next section. These comparative-statics results can be derived from the price equations (14)–(16) and from (A.2), which characterizes the steady-state investor masses.

3. MONOPOLISTIC MARKETMAKING

We assume here that investors can trade with the monopolistic marketmaker only when they meet one of the marketmaker's nonatomic "dealers." There is a unit mass of such dealers who contact potential investors randomly and pairwise independently, letting ρ be the intensity with which a dealer contacts a given agent. Dealers instantly balance their positions with their marketmaking firm, which, on the whole, holds no inventory.

With these assumptions, the equilibrium is computed as in Section 2. The masses are determined by (3)–(6) and the prices are given by Theorem 2.

It might seem surprising that a single monopolistic marketmaker is equivalent for pricing purposes to many "competing" nonatomic marketmakers. The result follows from the fact that a search economy is inherently uncompetitive, in that each time agents meet, a bilateral bargaining relationship obtains. With many nonatomic marketmakers, however, it is natural to assume that $z < 1$, while a monopolistic marketmaker could be assumed to have all of the bargaining power ($z = 1$). In practice, monopolists could develop dominant bargaining power by building a reputation for being "tough," or by being larger and wealthier than small investors.[16]

For these reasons, the label "monopolistic" serves to separate the case $z = 1$ from the case $z < 1$. The distinction between monopolistic and competitive marketmakers is clarified when search intensities are endogenized in Section 7.

A monopolistic marketmaker quotes an ask price A and a bid price B that are, respectively, a buyer's and a seller's reservation value. Hence, in equilibrium, $B \leq P \leq A$.

[16] In our model, a marketmaker's profit is not affected by any one infinitesimal trade. Further, Board and Zwiebel (2003) show that if agents bid resources for the right to make an offer, one agent much richer than another endogenously receives the entire bargaining power.

4. FAST SEARCH LEADS TO COMPETITIVE PRICES?

A competitive Walrasian equilibrium is characterized by a single price process at which agents may buy and sell *instantly*, such that supply equals demand in each state and at every point in time. A Walrasian allocation is efficient and all assets are held by agents of high type, if there are enough such agents,[17] which is the case in steady state if $s < \lambda_u/(\lambda_u + \lambda_d)$. In this case, the unique Walras equilibrium has agent masses

$$\begin{aligned} &\mu^*_{ho} = s, \\ &\mu^*_{hn} = \frac{\lambda_u}{\lambda_u + \lambda_d} - s, \\ &\mu^*_{lo} = 0, \\ &\mu^*_{ln} = \frac{\lambda_d}{\lambda_u + \lambda_d}, \end{aligned} \tag{17}$$

and price

$$P^* = E_t\left[\int_t^\infty e^{-r(s-t)}\,ds\right] = \frac{1}{r}, \tag{18}$$

which may be viewed as the reservation value of holding the asset forever for a hypothetical investor who is always of high type.

In the case that $s > \lambda_u/(\lambda_u + \lambda_d)$, the masses are determined similarly, and since the marginal investor has low liquidity, the Walrasian price is the reservation value of holding the asset indefinitely for a hypothetical agent who is permanently of low type (that is, $P^* = (1 - \delta)/r$). If $s = \lambda_u/(\lambda_u + \lambda_d)$, then any price P^* between $1/r$ and $(1 - \delta)/r$ is a Walrasian equilibrium.

Faster search by either investors or marketmakers leads in the limit to the efficient allocations μ^* of the Walrasian market. The following theorem further determines the circumstances under which prices approach the competitive Walrasian prices, P^*.

THEOREM 3: *Let* $(\lambda^k, \rho^k, \mu^k, B^k, A^k, P^k)$ *be a sequence of stationary equilibria.*

1. Fast Investors. *If* $\lambda^k \to \infty$, (ρ^k) *is any sequence, and* $0 < q < 1$, *then* $\mu^k \to \mu^*$, *and* B^k, A^k, *and* P^k *converge to a Walrasian price* P^*.

2. Fast Competing Marketmakers. *If* $\rho^k \to \infty$, (λ^k) *is any sequence, and* $z < 1$, *then* $\mu^k \to \mu^*$, *and* B^k, A^k, *and* P^k *converge to a Walrasian price* P^*.

[17] The quantity of such agents can be thought, for instance, as the capacity for taking a certain kind of risk.

3. Fast Monopolistic Marketmaker. *If $\lambda^k = \lambda$ is constant, $\rho^k \to \infty$ is an increasing sequence, and $z = 1$, then $\mu^k \to \mu^*$ and the bid–ask spread, $A^k - B^k$, is increasing.*

Part 1 shows that prices become competitive and that the bid–ask spread approaches zero as investors find *each other* more quickly, regardless of the nature of intermediation. In other words, the availability to investors of a search alternative forces marketmakers to offer relatively competitive prices, consistent with the evidence of Lamoureux and Schnitzlein (1997).[18]

Part 2 shows that fast intermediation by competing marketmakers also leads to competitive prices and vanishing bid–ask spreads. This may seem surprising, given that an investor trades with the first encountered marketmaker, and this marketmaker could have almost all bargaining power (z close to 1). As ρ increases, however, the investor's outside option when bargaining with a marketmaker improves, because he can more easily meet another marketmaker, and this sequential competition ultimately results in competitive prices.

Part 3 shows that fast intermediation by a monopolistic marketmaker does not lead to competitive prices. In fact, the bid–ask spread *widens* as intermediation by marketmakers increases. This is because an investor's potential "threat" to search for a direct trade with another investor becomes increasingly less persuasive, since the mass of investors with whom there are gains from trade is shrinking.

Contrary to our result, Rubinstein and Wolinsky (1985) find that their bargaining equilibrium (without intermediaries) does *not* converge to the competitive equilibrium as trading frictions approach zero. Gale (1987) argues that this failure is due to the fact that the total mass of agents entering their economy is infinite, which makes the competitive equilibrium of the total economy undefined. Gale (1987) shows that if the total mass of agents is finite, then the economy (which is not stationary) is Walrasian in the limit. He suggests that, when considering stationary economies, one should compare the bargaining prices to those of a "flow equilibrium" rather than a "stock equilibrium." Our model has a natural determination of steady-state masses, even though no agent enters the economy. This is accomplished by considering agents whose types change over time.[19] We are able to reconcile a steady-state economy with convergence to Walrasian outcomes in both a flow and stock sense, both for

[18] This result holds, under certain conditions, even if the monopolistic marketmaker can be approached instantly ($\rho = +\infty$). In this case, for any finite λ, *all* trades are done using the marketmaker, but as the investors' outside options improve, even a monopolistic marketmaker needs to quote competitive prices.

[19] Gale (1986a, 1986b) and McLennan and Sonnenschein (1991) show that a bargaining game implements Walrasian outcomes in the limiting case with no frictions (that is, no discounting) in much richer settings for preferences and goods. See also Binmore and Herrero (1988).

allocations and for prices, and by increasing both investor search and market-maker search.[20]

5. NUMERICAL EXAMPLE

We illustrate some of the search effects on asset pricing and marketmaking with a numerical example. Figure 1 shows the marketmakers' bid (B) and ask (A) prices as well as the interinvestor price (P). These prices are plotted as functions of the intensity, ρ, of meeting dealers. The top panel deals with the case of competing marketmakers with bargaining power $z = 0.8$, whereas the bottom panel treats a monopolistic marketmaker ($z = 1$). The parameters that underlie these graphs are as follows. First, $\lambda_d = 0.1$ and $\lambda_u = 1$, which implies that an agent is of high liquidity type 91% of the time. An investor finds other investors on average every two weeks, that is, $\lambda = 26$, and selling investors have bargaining power $q = 0.5$. The supply is $s = 0.8$ and the interest rate is $r = 0.05$.

Since allocations become more efficient as ρ increases, for both the competitive and monopolistic cases, all prices increase with ρ. Interestingly, in the case of competing marketmakers ($z = 0.8$), the bid–ask spread decreases to zero and the prices increase to the Walrasian price $1/r = 20$. In the case of a monopolist marketmaker ($z = 1$), on the other hand, the bid–ask spread is increasing in ρ and, due to this spread, the prices are bounded away from $1/r = 20$.

The intuition for this difference is as follows. When the dealers' contact intensities increase, they execute more trades. Investors then find it more difficult to contact other investors with whom to trade. If dealers have all of the bargaining power, this leads to wider spreads. If dealers do not have all of the bargaining power, however, then higher marketmaker intensity leads to a narrowing of the spread, because an investor has an improved threat of waiting to trade with the next encountered marketmaker.

6. HETEROGENEOUS INVESTORS

So far, we have assumed that investors are homogeneous with respect to the speed with which they find counterparties. In certain OTC markets, however, some investors are more sophisticated than others, in the sense that they have faster and easier access to counterparties. To capture this effect, we assume that there are two different investor classes: "sophisticated," of total mass μ^s, and "unsophisticated," of mass $1 - \mu^s$. We assume that sophisticated investors meet marketmakers with an intensity ρ^s, while unsophisticated investors meet

[20] Other important differences between our framework and that of Rubinstein and Wolinsky (1985) are that we accommodate repeated trade and we diminish search frictions explicitly through λ rather than implicitly through the discount rate. See Bester (1988, 1989) for the importance of diminishing search frictions directly.

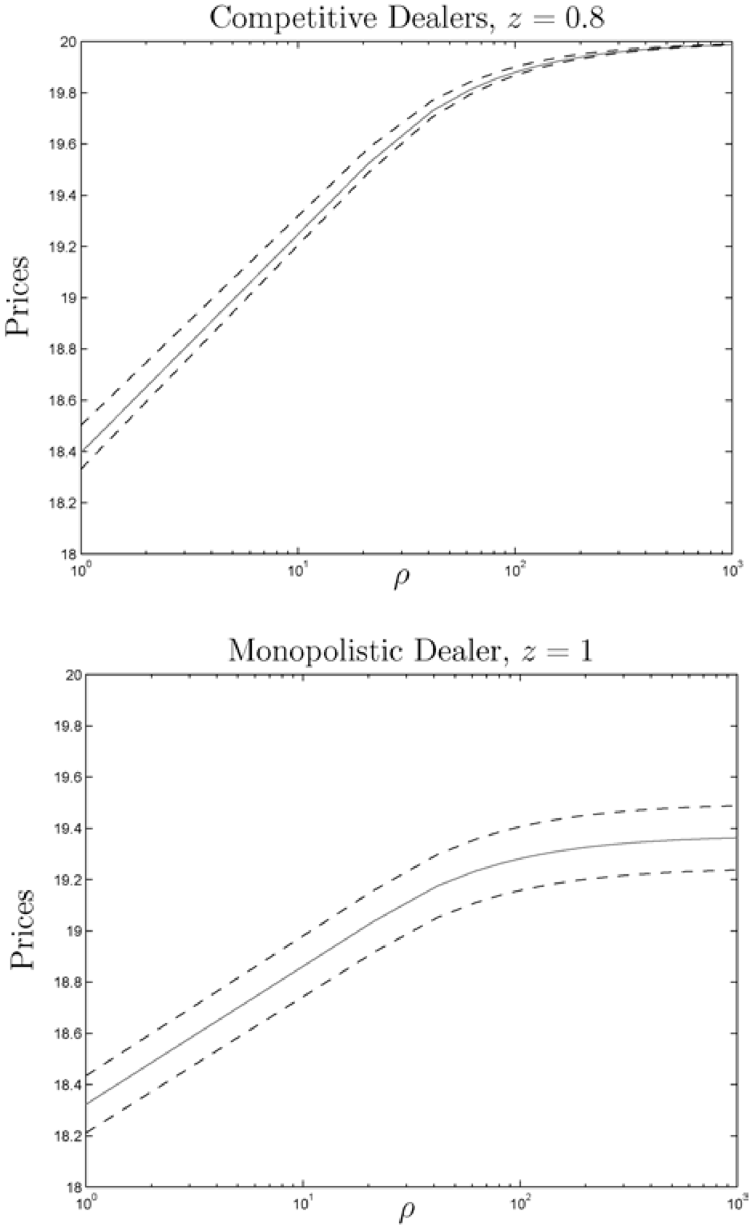

FIGURE 1.—The solid line shows the price P for trades between investors; the dashed lines show the bid (B) and ask (A) prices applied by marketmakers. The prices are functions of the intensity (ρ) with which an investor meets a dealer, which is plotted on a logarithmic scale. The bargaining power z of the marketmaker is 0.8 in the top panel and 1 in the bottom panel.

marketmakers at intensity ρ^u, where $\rho^u < \rho^s$. We assume here that investors cannot trade directly with each other, that is, $\lambda = 0$. If this assumption is relaxed and investors are able to find each other (possibly with type-dependent speeds), then the nature of the equilibrium that we will describe would change for certain parameters. In particular, sophisticated investors would, under certain conditions, profit from executing as many trades as possible and would start acting like marketmakers. This interesting effect is beyond the scope of this paper; we focus on how marketmakers react to differences in investor sophistication.

An investor's type is observable to marketmakers, who have bargaining power $z < 1$. When a sophisticated investor meets a marketmaker, the outcome of their bargaining is a bid price of B^s or an ask price of A^s. An unsophisticated investor takes more time to locate a marketmaker, resulting in higher expected holding costs and a poorer bargaining position. Hence, unsophisticated investors receive different bid and ask prices, which we denote by B^u and A^u, respectively.

When the supply of shares is so low that sophisticated investors are "marginal" buyers, then all unsophisticated investors optimally stay out of the market, that is, they never buy. Similarly, when the asset supply is large, sophisticated investors are marginal sellers, and unsophisticated investors hold the asset, never selling. With an intermediate supply, all investors trade, but unsophisticated investors trade at a larger spread. The following theorem characterizes the most important properties of the equilibrium with heterogeneous investors; a full characterization is in the Appendix.

THEOREM 4: *In equilibrium, unsophisticated investors do not trade if* $s < \mu^s(\lambda_u/(\lambda_u + \lambda_d))$ *or* $s > 1 - \mu^s(\lambda_d/(\lambda_u + \lambda_d))$. *Otherwise, all investors trade and marketmakers quote a larger bid–ask spread to unsophisticated investors than to sophisticated investors. That is,* $A^u - B^u > A^s - B^s$. *In particular, an agent who meets a marketmaker with intensity* ρ *faces a bid–ask of*

$$A - B = \frac{z\delta}{r + \lambda_u + \lambda_d + \rho(1 - z)}. \tag{19}$$

7. ENDOGENOUS SEARCH AND WELFARE

Here, we investigate the search intensities that marketmakers would optimally choose in the two cases considered above: a single monopolistic marketmaker and nonatomic competing marketmakers. We illustrate how marketmakers' choices of search intensities depend on (i) a marketmaker's personal influence on the equilibrium allocations of assets and (ii) a marketmaker's bargaining power. We take investors' search intensity λ as given, and assume that the meeting intensity ρ between investors and marketmakers is determined solely by marketmakers' technology choice. Considering the interactions that arise if both investors and intermediaries choose search intensities would be an interesting issue for future research.[21]

Because the marketmakers' search intensities, collectively, affect the masses μ of investor types, it is natural to take as given the initial masses, $\mu(0)$, of investors, rather than to compare based on the different steady-state masses that correspond to different choices of search intensities. Hence, in this section, we are not relying on a steady-state analysis.

We assume that a marketmaker chooses one search intensity and abides by it. This assumption is convenient and can be motivated by interpreting the search intensity as based on a technology that is difficult to change. A full dynamic analysis of the optimal control of marketmaking intensities with small switching costs would be interesting, but seems difficult. We merely assume that marketmakers choose ρ so as to maximize the present value, using their discount rate r, of future marketmaking spreads, net of the rate $\Gamma(\rho)$ of technology costs, where $\Gamma: [0, \infty) \to [0, \infty)$ is assumed for technical convenience to be continuously differentiable, strictly convex, with $\Gamma(0) = 0$, $\Gamma'(0) = 0$, and $\lim_{\rho\to\infty} \Gamma'(\rho) = \infty$.

The marketmaker's trading profit, per unit of time, is the product of the volume of trade, $\rho\mu_m$, and the bid–ask spread, $A - B$. Hence, a monopolistic marketmaker who searches with an intensity of ρ has an initial valuation of

$$\pi^M(\rho) = E\left[\int_0^\infty \rho\mu_m(t,\rho)(A(t,\rho) - B(t,\rho))e^{-rt}\,dt\right] - \frac{\Gamma(\rho)}{r}, \tag{20}$$

where $\mu_m = \min\{\mu_{lo}, \mu_{hn}\}$ and where we are using the obvious notation to indicate dependence of the solution on ρ and t.

Any one nonatomic marketmaker does not influence the equilibrium masses of investors and, therefore, values his profits at

$$\pi^C(\rho) = \rho E\left[\int_0^\infty \mu_m(t)(A(t) - B(t))e^{-rt}\,dt\right] - \frac{\Gamma(\rho)}{r}.$$

An equilibrium intensity, ρ^C, for nonatomic marketmakers is a solution to the first-order condition

$$\Gamma'(\rho^C) = rE\left[\int_0^\infty \mu_m(t,\rho^C)\big(A(t,\rho^C) - B(t,\rho^C)\big)e^{-rt}\,dt\right]. \tag{21}$$

The following theorem characterizes equilibrium search intensities in the case of "patient" marketmakers.

[21] Related to this, Pagano (1989) considers a one-period model in which investors choose between searching for a counterparty and trading on a centralized market.

THEOREM 5: *There exists a marketmaking intensity ρ^M that maximizes $\pi^M(\rho)$. There exists $\bar{r} > 0$ such that, for all $r < \bar{r}$ and for each $z \in [0, 1]$, a unique number $\rho^C(z)$ solves the optimal search intensity condition* (21). *Moreover*, $\rho^C(0) = 0$, *$\rho^C(z)$ is increasing in z, and $\rho^C(1)$ is larger than any solution*, ρ^M, *to the monopolist's problem.*[22]

In addition to providing the existence of equilibrium search intensities, this result establishes that (i) competing marketmakers provide more marketmaking services if they can capture a higher proportion of the gains from trade and (ii) competing marketmakers with full bargaining power provide more marketmaking services than a monopolistic marketmaker, since they do not internalize the consequences of their search on the masses of investor types.

To consider the welfare implications of marketmaking in our search economy, we adopt as a notion of "social welfare" the sum of the utilities of investors and marketmakers. This can be interpreted as the total investor utility in the case in which the marketmaker profits are redistributed to investors, for instance, through asset holdings. With our form of linear preferences, maximizing social welfare is a meaningful concept in that it is equivalent to requiring that utilities cannot be Pareto improved by changing allocations and by making initial consumption transfers.[23] By "investor welfare," we mean the total of investors' utilities, assuming that marketmakers' profits are not redistributed to investors. We take "marketmaker welfare" to be the total valuation of marketmaking profits, net of the cost of intermediation.

In our risk-neutral framework, welfare losses are easily quantified. The total "social loss" is the cost $\Gamma(\rho)$ of intermediation plus the present value of the stream $\delta\mu_{lo}(t), t \geq 0$, of dividends wasted through misallocation. At a given marketmaking intensity ρ, this leaves the social welfare

$$w^S(\rho) = E\left[\int_0^\infty (s - \delta\mu_{lo}(t))e^{-rt}\,dt\right] - \frac{\Gamma(\rho)}{r}.$$

Investor welfare is, similarly,

$$\begin{aligned} w^I(\rho) = E\bigg[\int_0^\infty \big(s &- \delta\mu_{lo}(t, \rho) \\ &- \rho\mu_m(t, \rho)(A(t, \rho) - B(t, \rho))\big)e^{-rt}\,dt\bigg] \end{aligned}$$

and the marketmakers' welfare is

$$w^M(\rho) = E\left[\int_0^\infty \rho\mu_m(t, \rho)(A(t, \rho) - B(t, \rho))e^{-rt}\,dt\right] - \frac{\Gamma(\rho)}{r}.$$

[22] If the monopolist's bargaining power is $z < 1$, it still holds that $\rho^C(z) > \rho^M(z)$.

[23] This "utilitarian" social welfare function can be justified by considering the utility of an agent "behind the veil of ignorance," not knowing what type of agent he will become.

We consider first the case of monopolistic marketmaking. We let ρ^M be the level of intermediation optimally chosen by the marketmaker and let ρ^S be the socially optimal level of intermediation. The relationship between the monopolistic marketmaker's chosen level ρ^M of intensity and the socially optimal intensity ρ^S is characterized in the following theorem.

THEOREM 6: *Let* $z = 1$. (i) *If investors cannot meet directly*, *that is*, $\lambda = 0$, *then the investor welfare* $w^I(\rho)$ *is independent of* ρ *and a monopolistic marketmaker provides the socially optimal level* ρ^S *of intermediation*; *that is*, $\rho^M = \rho^S$. (ii) *If* $\lambda > 0$, *then provided* q *is* 0 *or* 1, $w^I(\rho)$ *decreases in* ρ *and the monopolistic marketmaker overinvests in intermediation*; *that is*, $\rho^M > \rho^S$.

The point of this result is that if investors cannot search, then their utilities do not depend on the level of intermediation because the monopolist extracts all gains from trade. In this case, because the monopolist gets all social benefits from providing intermediation and bears all of the costs, he chooses the socially optimal level of intermediation.

If, on the other hand, investors can trade directly with each other, then the marketmaker may exploit the opportunity to invest in additional search for trades so as to reduce the opportunities of investors to trade directly with each other. Therefore, investor welfare decreases with ρ. Consequently, the marketmaker's marginal benefit from intermediation is larger than the social benefit, so there is too much intermediation.[24]

We now turn to the case of nonatomic (competing) marketmakers. We saw above that the equilibrium level of intermediation of a nonatomic marketmaker depends critically on its bargaining power. With no bargaining power, such a marketmaker provides no intermediation. With complete bargaining power, they search more than a monopolistic marketmaker would.

A government may sometimes be able to affect intermediaries' market power, for instance through the enforcement of regulation (DeMarzo, Fishman, and Hagerty (2000)). Hence, we consider the following questions: How much marketmaker market power is socially optimal? How much market power would the intermediaries choose to have? Would investors prefer that marketmakers have some market power? These questions are answered in the following theorem, in which we let z^I, z^S, and z^M denote the marketmaker bargaining power that would be chosen by, respectively, the investors, a social-welfare maximizing planner, and marketmakers.

THEOREM 7: *It holds that* $z^I > 0$. *There is some* $\bar{r} > 0$ *such that*, *provided* $r < \bar{r}$, *we have* $z^I < z^S \leq z^M = 1$.

[24] If $0 < q < 1$, then increasing ρ has the additional effect of changing the relative strength of investors' bargaining positions with the marketmaker, because it changes their outside options, which complicates the calculations.

Investors in our model would prefer to enter a market in which nonatomic marketmakers have some market power $z^I > 0$, because this gives marketmakers an incentive to provide intermediation. The efficient level of intermediation is achieved with even higher marketmakers, power $z^S > z^I$. Marketmakers themselves prefer to have full bargaining power.

8. EMPIRICAL IMPLICATIONS

This paper lays out a theory of asset pricing and marketmaking based on search and bargaining. We show how search-based inefficiencies affect prices through equilibrium allocations and through the effect of search on agents' bargaining positions, that is, their outside options based on their ability to trade with other investors or marketmakers.

Consider, for example, the OTC market for interest-rate swaps, which, according to the British Bankers Association has open positions totalling roughly \$100 trillion dollars. Customers rarely have material private information about the current level of the underlying interest rates, so standard information-based explanations of bid–ask spreads are not compelling in this market. Instead, a "sales trader" sets spreads based on a customer's (perceived) outside option and would rarely fear that the customer has superior information about the underlying interest rates. The customer's outside option depends on how easily he can find a counterparty himself (proxied by λ in our model), and how easily he can access other banks (proxied by ρ in our model). To trade OTC derivatives with a bank, one needs, among other things, an account and a credit clearance. Smaller investors often have an account with only one or a few banks, lowering their search options. Hence, a testable implication of our search framework is that smaller investors, typically those with fewer search options, receive less favorable prices. We note that these investors are less likely to be informed, so traditional information-based models of spreads (for example, Glosten and Milgrom (1985)), applied to this market, would have the opposite prediction. Consistent with our results, Schultz (2001) finds that bid–ask spreads are larger for smaller trades and for smaller (institutional) investors in the market for corporate bonds. Furthermore, Green, Hollifield, and Schurhoff (2004) and Harris and Piwowar (2004) find that bid–ask spreads are larger for smaller trades and for more complex instruments in the market for municipal bonds.

The model that we present here can also be viewed as one of imperfect competition, for example, in specialist-based equity markets. In particular, the model shows that even a monopolistic marketmaker may have a tight bid–ask spread if investors can easily trade directly with each other (that is, have a high λ). This resembles situations at the New York Stock Exchange (NYSE) in which, despite a single specialist for each stock, floor brokers can find and trade among themselves, and outside brokers can find each other and trade "around" the specialist with limit orders. However, on Nasdaq, a "phone market" with several dealers for each stock, it can be difficult for investors to find each other directly. Before the reforms of 1994, 1995, and 1997, it was difficult for investors to compete with Nasdaq marketmakers through limit orders.[25] This may help explain why spreads were higher on Nasdaq than on NYSE (Huang and Stoll (1996)). Consistent with this view, Barclay, Christie, Harris, Kandel, and Schultz (1999) find that the "Securities and Exchange Commission began implementing reforms that would permit the public to compete directly with Nasdaq dealers by submitting binding limit orders. ... Our results indicate that quoted and effective spreads fell dramatically."

The competition faced by marketmakers from direct trade between investors can perhaps be gauged by the *participation rate* of marketmakers, that is, the fraction of trades that are intermediated by a marketmaker. Our model suggests that, with equal marketmaker availability and stock characteristics, stocks with higher participation rates are characterized by lower search intensity (λ) and, hence, higher bid–ask spreads. On Nasdaq, the participation rate was once large relative to the NYSE, whose participation rate was between 18.8% and 24.2% in the 1990s (New York Stock Exchange (2001)). At that time, the NYSE may well have covered stocks whose investors had higher direct contact rates (λ) than those covered, on average, by Nasdaq.

Our modeled counterparty search times can proxy, in practice, also for delays necessary for counterparties to verify one another's credit standing, and to arrange for trade authorization and financing or for the time necessary to familiarize an investor with a product type or contractual terms. Even in an OTC market as liquid as that of U.S. Treasuries, delays necessary to contact suitable counterparties are frequently responsible for meaningful price effects, for example, as documented by Krishnamurthy (2002). Duffie, Gârleanu, and Pedersen (2003) provide additional discussion of the empirical relevance of search for asset pricing behavior.

Graduate School of Business, Stanford University, Stanford, CA 94305-5015, U.S.A.; duffie@stanford.edu,

Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104-6367, U.S.A.; garleanu@wharton.upenn.edu,

and

Stern School of Business, New York University, 44 West Fourth Street, Suite 9-190, New York, NY 10012-1126, U.S.A; lpederse@stern.nyu.edu.

Manuscript received April, 2004; final revision received February, 2005.

[25] See Barclay, Christie, Harris, Kandel, and Schultz (1999) and references therein.

APPENDIX: PROOFS

PROOF OF PROPOSITION 1: Start by letting

$$y=\frac{\lambda_u}{\lambda_u+\lambda_d},$$

and assume that $y>s$. The case $y\leq s$ can be treated analogously. Setting the right-hand side of (3) to zero and substituting all components of μ other than μ_{lo} in terms of μ_{lo} from (1) and (2) and from $\mu_{lo}+\mu_{ln}=\lambda_d(\lambda_d+\lambda_u)^{-1}=1-y$, we obtain the quadratic equation

$$Q(\mu_{lo})=0, \tag{A.1}$$

where

$$Q(x)=2\lambda x^2+\big(2\lambda(y-s)+\rho+\lambda_u+\lambda_d\big)x-\lambda_d s. \tag{A.2}$$

It is immediate that Q has a negative root (since $Q(0)<0$) and has a root in the interval $(0,1)$ (since $Q(1)>0$).

Since μ_{lo} is the largest and positive root of a quadratic with positive leading coefficient and with a negative root, to show that $\mu_{lo}<\eta$ for some $\eta>0$, it suffices to show that $Q(\eta)>0$. Thus, so that $\mu_{ho}>0$ (for, clearly, $\mu_{ho}<1$), it is sufficient that $Q(s)>0$, which is true, since

$$Q(s)=2\lambda s^2+(\lambda_u+2\lambda(y-s)+\rho)s.$$

Similarly, $\mu_{ln}>0$ if $Q(1-y)>0$, which holds because

$$Q(1-y)=2\lambda(1-y)^2+(2\lambda(y-s)+\rho)(1-y)+\lambda_d(1-s).$$

Finally, since $\mu_{hn}=y-s+\mu_{lo}$, it is immediate that $\mu_{hn}>0$.

We present a sketch of a proof of the claim that from any admissible initial condition $\mu(0)$, the system converges to the steady state μ.

Because of the two restrictions (1) and (2), the system is reduced to two equations, which can be thought of as equations in the unknowns $\mu_{lo}(t)$ and $\mu_l(t)$, where $\mu_l(t)=\mu_{lo}(t)+\mu_{ln}(t)$. The equation for $\mu_l(t)$ does not depend on $\mu_{lo}(t)$, and admits the simple solution

$$\mu_l(t)=\mu_l(0)e^{-(\lambda_d+\lambda_u)t}+\frac{\lambda_d}{(\lambda_d+\lambda_u)}\big(1-e^{-(\lambda_d+\lambda_u)t}\big).$$

Define the function

$$\begin{aligned}G(w,x)={}&-2\lambda x^2-\big(\lambda_u+\lambda_d+2\lambda(1-s-w)+\rho\big)x\\&+\rho\max\{0,s+w-1\}+\lambda_d s\end{aligned}$$

and note that μ_{lo} satisfies

$$\dot{\mu}_{lo}(t) = G(\mu_l(t), \mu_{lo}(t)).$$

The claim is proved by the following steps:

1. Choose t_1 high enough that $s + \mu_l(t) - 1$ does not change sign for $t > t_1$.

2. Show that $\mu_{lo}(t)$ stays in $(0, 1)$ for all t by verifying that $G(w, 0) > 0$ and $G(w, 1) < 0$.

3. Choose t_2 $(\geq t_1)$ high enough that $\mu_l(t)$ changes by at most an arbitrarily chosen $\varepsilon > 0$ for $t > t_2$.

4. Note that, for any value $\mu_{lo}(t_2) \in (0, 1)$, the equation

$$\dot{x}(t) = G(w, x(t)) \tag{A.3}$$

with boundary condition $x(t_2) = \mu_{lo}(t_2)$ admits a solution that converges exponentially, as $t \to \infty$, to a positive quantity that can be written as $(-b + \sqrt{b^2 + c})$, where b and c are affine functions of w. The convergence is uniform in $\mu_{lo}(t_2)$.

5. Finally, using a comparison theorem (for instance, see Birkhoff and Rota (1969, p. 25)), $\mu_{lo}(t)$ is bounded by the solutions to (A.3) that correspond to w taking the highest and lowest values of $\mu_l(t)$ for $t > t_2$ (these are, of course, $\mu_l(t_2)$ and $\lim_{t\to\infty} \mu_l(t)$). By virtue of the previous step, for high enough t, these solutions are within $O(\varepsilon)$ of the steady-state solution μ_{lo}. *Q.E.D.*

PROOF OF THEOREM 2: To calculate V_σ and P, we consider a particular agent and a particular time t, let τ_l denote the next (stopping) time at which that agent's intrinsic type changes, let τ_i denote the next (stopping) time at which another investor with gain from trade is met, let τ_m denote the next time a marketmaker is met, and let $\tau = \min\{\tau_l, \tau_i, \tau_m\}$. Then,

$$\begin{aligned}
V_{lo} &= E_t\biggl[\int_t^{\tau} e^{-r(u-t)}(1-\delta)\,du + e^{-r(\tau_l-t)}V_{ho}\mathbb{1}_{\{\tau_l=\tau\}} \\
&\qquad + e^{-r(\tau_i-t)}(V_{ln}+P)\mathbb{1}_{\{\tau_i=\tau\}} + e^{-r(\tau_m-t)}(V_{ln}+B)\mathbb{1}_{\{\tau_m=\tau\}}\biggr], \\
V_{ln} &= E_t\bigl[e^{-r(\tau_l-t)}V_{hn}\bigr], \\
V_{ho} &= E_t\biggl[\int_t^{\tau_l} e^{-r(u-t)}\,du + e^{-r(\tau_l-t)}V_{lo}\biggr], \\
V_{hn} &= E_t\bigl[e^{-r(\tau_l-t)}V_{ln}\mathbb{1}_{\{\tau_l=\tau\}} + e^{-r(\tau_i-t)}(V_{ho}-P)\mathbb{1}_{\{\tau_i=\tau\}}, \\
&\qquad + e^{-r(\tau_m-t)}(V_{ho}-A)\mathbb{1}_{\{\tau_m=\tau\}}\bigr],
\end{aligned} \tag{A.4}$$

where E_t denotes expectation conditional on the information available at time t. Differentiating both sides of (A.4) with respect to t, we get (10).

In steady state, $\dot{V}_\sigma = 0$ and hence (10) implies the following equations for the value functions and prices:

$$
\begin{aligned}
\text{(A.5)} \qquad V_{lo} &= \frac{(\lambda_u V_{ho} + 2\lambda\mu_{hn} P + \rho B + (2\lambda\mu_{hn} + \rho) V_{ln} + 1 - \delta)}{r + \lambda_u + 2\lambda\mu_{hn} + \rho}, \\
V_{ln} &= \frac{\lambda_u V_{hn}}{r + \lambda_u}, \\
V_{ho} &= \frac{\lambda_d V_{lo} + 1}{r + \lambda_d}, \\
V_{hn} &= \frac{(\lambda_d V_{ln} + (2\lambda\mu_{lo} + \rho) V_{ho} - 2\lambda\mu_{lo} P - \rho A)}{r + \lambda_d + 2\lambda\mu_{lo} + \rho}.
\end{aligned}
$$

(We note that agents on the "long side" of the market are rationed when they interact with the marketmaker and, therefore, their trading intensity with the marketmaker is less than ρ. This does not affect (A.5), however, because the price is the reservation value.) Define $\Delta V_l = V_{lo} - V_{ln}$ and $\Delta V_h = V_{ho} - V_{hn}$ to be the reservation values. With this notation, the prices are determined by

$$
\begin{aligned}
\text{(A.6)} \qquad P &= \Delta V_l (1 - q) + \Delta V_h q, \\
A &= \Delta V_h z + M(1 - z), \\
B &= \Delta V_l z + M(1 - z), \\
M &= \begin{cases} \Delta V_h, & \text{if } s < \dfrac{\lambda_u}{\lambda_u + \lambda_d}, \\ \Delta V_l, & \text{if } s > \dfrac{\lambda_u}{\lambda_u + \lambda_d}, \end{cases}
\end{aligned}
$$

and $M \in [\Delta V_l, \Delta V_h]$ if $s = \lambda_u / (\lambda_u + \lambda_d)$. Let

$$
\begin{aligned}
\psi_d &= \lambda_d + 2\lambda\mu_{lo}(1 - q) + (1 - \tilde{q})\rho(1 - z), \\
\psi_u &= \lambda_u + 2\lambda\mu_{hn} q + \tilde{q}\rho(1 - z),
\end{aligned}
$$

where

$$
\tilde{q} \begin{cases} = 1, & \text{if } s < \dfrac{\lambda_u}{\lambda_u + \lambda_d}, \\ = 0, & \text{if } s > \dfrac{\lambda_u}{\lambda_u + \lambda_d}, \\ \in [0, 1], & \text{if } s = \dfrac{\lambda_u}{\lambda_u + \lambda_d}. \end{cases}
$$

With this notation, we see that appropriate linear combinations of (A.5)–(A.6) yield

$$\begin{bmatrix} r+\psi_u & -\psi_u \\ -\psi_d & r+\psi_d \end{bmatrix} \begin{bmatrix} \Delta V_l \\ \Delta V_h \end{bmatrix} = \begin{bmatrix} 1-\delta \\ 1 \end{bmatrix}.$$

Consequently,

$$\begin{bmatrix} \Delta V_l \\ \Delta V_h \end{bmatrix} = \frac{1}{r}\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{\delta}{r}\frac{1}{r+\psi_u+\psi_d}\begin{bmatrix} r+\psi_d \\ \psi_d \end{bmatrix}, \tag{A.7}$$

which leads to the price formula stated by the theorem. Note also that $\Delta V_l < \Delta V_h$.

Finally, we need to verify that any agent prefers, at any time, given all information, to play the proposed equilibrium trading strategy, assuming that other agents do. It is enough to show that an agent agrees to trade at the candidate equilibrium prices when contacted by an investor with whom there are potential gains from trade.

The Bellman principle for an agent of type *lo* in contact with an agent of type *hn* is

$$\begin{aligned} &P + V_{ln} \\ &\quad \geq E_t\bigg[\int_t^\tau e^{-r(u-t)}(1-\delta)\,du + e^{-r(\tau_l-t)}V_{ho}\mathbb{1}_{\{\tau_l=\tau\}} \\ &\qquad + e^{-r(\tau_i-t)}(V_{ln}+P)\mathbb{1}_{\{\tau_i=\tau\}} + e^{-r(\tau_m-t)}(V_{ln}+B)\mathbb{1}_{\{\tau_m=\tau\}}\bigg], \end{aligned}$$

where $\tau = \min\{\tau_l, \tau_i, \tau_m\}$. This inequality follows from that fact that $\Delta V_h \geq P \geq \Delta V_l$, which means that selling the asset, consuming the price, and attaining the candidate value of a nonowner with low valuation, dominates (at least weakly) the value of keeping the asset, consuming its dividends, and collecting the discounted expected candidate value achieved at the next time τ_m of a trading opportunity or at the next time τ_r of a change to a low discount rate, whichever comes first. There is a like Bellman inequality for an agent of type *hn*.

Now, to verify the sufficiency of the Bellman equations for individual optimality, consider any initial agent type $\sigma(0)$ and any feasible trading strategy θ, an adapted process whose value is 1 whenever the agent owns the asset and 0 whenever the agent does not own the asset. The associated type process σ^θ and a wealth process of $W = 0$ (which can be assumed without loss of generality) determine a cumulative consumption process C^θ that satisfies

$$dC_t^\theta = \theta_t(1-\delta\mathbb{1}_{\{\sigma^\theta(t)=lo\}})\,dt - \hat{P}\,d\theta_t. \tag{A.8}$$

Following the usual verification argument for stochastic control, for any future meeting time τ^m, $m \in \mathbb{N}$, we have

$$V_{\sigma(0)} \geq E\left[\int_0^{\tau^m} e^{-rt}\, dC_t^\theta\right] + E\big[e^{-r\tau^m} V_{\sigma^\theta(\tau^m)}\big].$$

(This assumes without loss of generality that a potential trading contact does not occur at time 0.) Letting m go to ∞, we have $V_{\sigma(0)} \geq U(C^\theta)$. Because $V_{\sigma(0)} = U(C^*)$, where C^* is the consumption process associated with the candidate equilibrium strategy, we have shown optimality. *Q.E.D.*

PROOF OF THEOREM 3: The convergence of the masses μ to μ^* is easily seen using (A.1), whether λ or ρ tends to infinity. Let us concentrate on the prices.

1. If $s < \lambda_u/(\lambda_u + \lambda_d)$, then we see using (A.1) that $\lambda\mu_{hn}$ tends to infinity with λ, while $\lambda\mu_{lo}$ is bounded. Hence, (A.7) shows that both ΔV_l and ΔV_h tend to r^{-1}, provided that $q > 0$. If $s > \lambda_u/(\lambda_u + \lambda_d)$, $\lambda\mu_{lo}$ tends to infinity with λ, while $\lambda\mu_{hn}$ is bounded. Hence, ΔV_l and ΔV_h tend to $r^{-1}(1-\delta)$, provided that $q < 1$. If $s = \lambda_u/(\lambda_u + \lambda_d)$, then $\lambda\mu_{hn} = \lambda\mu_{lo}$ tends to infinity with λ, and ΔV_l and ΔV_h tend to $r^{-1}(1-\delta(1-q))$. In each case, the reservation values converge to the same value, which is a Walrasian price.
2. Equation (A.7) shows that both ΔV_l and ΔV_h tend to the Walrasian price $r^{-1}(1-\delta(1-\tilde{q}))$ as ρ approaches infinity.
3. When $z = 1$, $A^k - B^k$ increases with ρ because $A - B = \delta(r + \psi_u + \psi_d)^{-1}$ and both ψ_u and ψ_d decrease, since μ_{lo} and μ_{hn} do. *Q.E.D.*

PROOF OF THEOREM 4: Let the value function of a sophisticated type-σ investor be V_σ^s and let the value function of an unsophisticated type-σ investor be V_σ^u. These value functions and the prices (A^s, B^s, A^u, B^u) are computed as in (A.5) and (A.6), with the modification that the interdealer price M is different. For any fixed interdealer price M, an agent who meets the marketmaker with intensity ρ, and who sells as a *lo* type and buys as an *hn* type (i.e., with $\Delta V_l \leq M \leq \Delta V_h$) has value functions determined by

$$\begin{aligned}
V_{ho}(r + \lambda_d) &= 1 + \lambda_d V_{lo},\\
V_{hn}(r + \lambda_d + \rho) &= \lambda_d V_{ln} + \rho\big(V_{ho} - [z\Delta V_h + (1-z)M]\big),\\
V_{ln}(r + \lambda_u) &= \lambda_u V_{hn},\\
V_{lo}(r + \lambda_u + \rho) &= 1 - \delta + \lambda_u V_{ho} + \rho\big(V_{ln} + [z\Delta V_l + (1-z)M]\big).
\end{aligned}$$

The system reduces to

$$\begin{aligned}
\Delta V_h(r + \lambda_d + \rho(1-z)) &= 1 + \lambda_d \Delta V_l + \rho(1-z)M,\\
\Delta V_l(r + \lambda_u + \rho(1-z)) &= 1 - \delta + \lambda_u \Delta V_h + \rho(1-z)M,
\end{aligned}$$

which implies that

$$\text{(A.9)}\quad \begin{bmatrix}\Delta V_l\\ \Delta V_h\end{bmatrix} = \frac{1+\rho(1-z)M}{r+\rho(1-z)}\begin{bmatrix}1\\1\end{bmatrix} - \frac{\delta}{r+\rho(1-z)}\frac{1}{r+\lambda_u+\lambda_d+\rho(1-z)} \times \begin{bmatrix} r+\lambda_d+\rho(1-z)\\ \lambda_d\end{bmatrix}.$$

Hence, this agent faces a bid–ask spread of

$$z(\Delta V_h - \Delta V_l) = \frac{z\delta}{r+\lambda_u+\lambda_d+\rho(1-z)}.$$

We show below, for each case, that M is given by

$$\text{(A.10)}\quad M = \begin{cases} \Delta V_h^s, & \text{if } s < \mu^s \dfrac{\lambda_u}{\lambda_u+\lambda_d},\\ \Delta V_h^u, & \text{if } \mu^s \dfrac{\lambda_u}{\lambda_u+\lambda_d} < s < \dfrac{\lambda_u}{\lambda_u+\lambda_d},\\ \Delta V_l^u, & \text{if } \dfrac{\lambda_u}{\lambda_u+\lambda_d} < s < 1-\mu^s\dfrac{\lambda_d}{\lambda_u+\lambda_d},\\ \Delta V_l^s, & \text{if } 1-\mu^s\dfrac{\lambda_d}{\lambda_u+\lambda_d} < s. \end{cases}$$

CASE (a): Consider first the case of $s < \mu^s\lambda_u/(\lambda_u+\lambda_d)$. The claim is that it is an equilibrium that the unsophisticated investors own no assets. Assuming this to be true, the market has only sophisticated investors, the interdealer price is $M = \Delta V_h^s$, and the buyers are rationed.

It remains to be shown that, with this interdealer price, there is no price at which marketmakers will sell and unsophisticated investors will buy. First, we note that the optimal response of an investor to the Markov (time-independent) investment problem can be chosen to be Markov, which means that one only needs to check the payoffs from Markov strategies that stipulate the same probability of trade for a given type at any time. The linearity of the problem further allows one to assume that the trading probability is 1 or 0. (When indifferent, the choice does not matter, so we may assume a corner solution.)

There are three possible Markov strategies for the unsophisticated investor that involve buying: buying as type h and selling as type l, buying as type l and selling as type h, and buying and holding (never selling).

If the unsophisticated investor buys as an h type and sells as an l type, then her value function satisfies (A.9), implying that $\Delta V_h^u < \Delta V_h^s = M$ since $\rho^u < \rho^s$.

The reservation values are even lower if she buys as an l and sells as an h type. Finally, if the unsophisticated investor buys and never sells, then her value function is also smaller than M. This is inconsistent with trading with the marketmaker, meaning that she never buys.

CASE (b): For the case $\mu_h^s < s < \mu_h$, the equilibrium is given by an interdealer price of $A^u = M = \Delta V_h^u = A(\rho^u)$. This is also the price at which unsophisticated hn agents buy from the marketmaker, and these agents are rationed. The sophisticated types hold a total $\mu_h^s = \mu^s \lambda_u / (\lambda_u + \lambda_d)$ of the supply, while the unsophisticated types hold the rest. This is clearly an equilibrium for the unsophisticated types. We must ensure that sophisticated types also behave optimally. In particular, we must check that $\Delta V_l^s \leq M \leq \Delta V_h^s$. For this, we use (A.7) and (A.9). We have $\Delta V_l^s \leq M$ if and only if

$$\frac{1 + \rho^s(1-z)M}{r + \rho^s(1-z)} - \frac{\delta(r + \lambda_d + \rho^s(1-z))}{r + \rho^s(1-z)} \frac{1}{r + \lambda_u + \lambda_d + \rho^s(1-z)} \leq M,$$

which holds if and only if

$$\frac{r + \lambda_d + \rho^s(1-z)}{r + \lambda_u + \lambda_d + \rho^s(1-z)} \geq \frac{\lambda_d}{r + \lambda_u + \lambda_d + \rho^u(1-z)},$$

which holds because $\rho^s \geq \rho^u$. Similarly, it can be verified that $M \leq \Delta V_h^s$ using the same formulae.

CASE (c): The remaining two cases are dual to those just treated. To see this, take the following new perspective of an agent's problem: An agent considers "acquiring" nonownership (that is, selling). The number of shares of nonownership is $1 - s$. If an l type acquires nonownership, then he gets a dividend of $-(1 - \delta)$ (that is, he gives up a dividend of $1 - \delta$). If an h type acquires nonownership, then he gets a dividend of -1. Said differently, he gets a dividend of $-(1 - \delta)$ like that of the l type and, additionally, he has a cost of δ. Hence, from this perspective, h and l types are reversed and the supply of shares is $1 - s$.

This explains why the equilibria in the latter two cases are mirror images of the equilibria in the former two cases. In particular, if $\lambda_u/(\lambda_u + \lambda_d) < s < 1 - \mu^s(\lambda_d/(\lambda_u + \lambda_d))$, then both sophisticated and unsophisticated investors trade, and the unsophisticated l type is rationed.

If $1 - \mu^s(\lambda_d/(\lambda_u + \lambda_d)) < s$, each unsophisticated investor owns a share and does not trade. (Using the alternative perspective, they are out of the market for nonownership.) The sophisticated investors hold the remaining $(1 - \mu^s)$ shares, they trade, and the selling sophisticated investors are rationed. *Q.E.D.*

PROOF OF THEOREM 5: There exists a number ρ^M that maximizes (19) since π^M is continuous and $\pi^M(\rho) \to -\infty$ as $\rho \to \infty$. We are looking for some $\rho^C \geq 0$ such that

$$\Gamma'(\rho^C) = rE \int_0^\infty \mu_m(\rho^C)\big(A(\rho^C) - B(\rho^C)\big)e^{-rt}\,dt. \tag{A.11}$$

Consider how both the left-hand and right-hand sides depend on ρ. The left-hand side is 0 for $\rho = 0$, increasing, and tends to infinity as ρ tends to infinity. For $z = 0$, $A(t, \rho) - B(t, \rho) = 0$ everywhere, so the right-hand side is zero, and, therefore, the unique solution to (A.11) is clearly $\rho^C = 0$. For $z > 1$, the right-hand side is strictly positive for $\rho = 0$. Furthermore, the steady-state value of the right-hand side can be seen to be decreasing, using the fact that μ_m is decreasing in ρ and using the explicit expression for the spread provided by (A.7). Furthermore, by continuity and still using (A.7), there is $\varepsilon > 0$ and T such that $\frac{\partial}{\partial\rho}\mu_m(A - B) < -\varepsilon$ for all $t > T$ and all r. Further, note that $t \mapsto re^{-rt}$ is a probability density function for any $r > 0$ and that the closer is r to zero, the more weight is given to high values of t (that is, the more important is the steady-state value for the integral). Therefore, the right-hand side is also decreasing in ρ for any initial condition on μ if r is small enough. These results yield the existence of a unique solution.

To see that $\rho^C > \rho^M$ when $z = 1$, consider the first-order conditions that determine ρ^M:

$$\begin{aligned} &\Gamma'(\rho^M) \\ &\quad = rE \int_0^\infty \Big[\mu_m(t, \rho^M)\big(A(t, \rho^M) - B(t, \rho^M)\big) \\ &\qquad\qquad + \rho^M \frac{\partial}{\partial \rho^M}\big(\mu_m(t, \rho^M)\big(A(t, \rho^M) - B(t, \rho^M)\big)\big)\Big]e^{-rt}\,dt. \end{aligned} \tag{A.12}$$

The integral of the first integrand term on the right-hand side of (A.12) is the same as that of (A.11), and that of the second is negative for small r. Hence, the right-hand side of (A.12) is smaller than the right-hand side of (A.11), implying that $\rho^C(1) > \rho^M$.

To see that $\rho^C(z)$ is increasing in z, we use the implicit function theorem and the dominated convergence theorem to compute the derivative of $\rho^C(z)$ with respect to z as

$$\frac{rE\int_0^\infty \mu_m(\rho^C)(A_z(\rho^C, z) - B_z(\rho^C, z))e^{-rt}\,dt}{\Gamma''(\rho^C) - rE\int_0^\infty (d/d\rho)\mu_m(\rho^C)(A(\rho^C, z) - B(\rho^C, z))e^{-rt}\,dt}. \tag{A.13}$$

If we use the steady-state expressions for μ, A, and B, this expression is seen to be positive because both the denominator and the numerator are positive. Hence, it is also positive with any initial masses if we choose r small enough. *Q.E.D.*

PROOF OF THEOREM 6: (i) The first part of the theorem, that the monopolistic marketmaker's search intensity does not affect investors when they can not search for each other, follows from (A.5), which shows that investor's utility is independent of ρ.

(ii) We want to prove that the investor welfare is decreasing in ρ, which directly implies that the marketmaker overinvests in intermediation services.

We introduce the notation $\Delta V_o = V_{ho} - V_{lo}$, $\Delta V_n = V_{hn} - V_{ln}$, and $\phi = \Delta V_h - \Delta V_l = \Delta V_o - \Delta V_n$, and start by proving a few general facts about the marketmaker spread, ϕ.

The dynamics of ϕ are given by the ordinary differential equation (ODE)

$$\dot{\phi}_t = \big(r + \lambda_d + \lambda_u + 2\lambda(1-q)\mu_{lo} + 2\lambda q \mu_{hn}\big)\phi_t - \delta.$$

Let $R = r + \lambda_d + \lambda_u + 2\lambda(1-q)\mu_{lo} + 2\lambda q\mu_{hn}$. The equation above readily implies that

$$\text{(A.14)}\qquad \frac{\partial \dot{\phi}_t}{\partial \rho} = R\frac{\partial \phi_t}{\partial \rho} + \left(2\lambda(1-q)\frac{\partial \mu_{lo}(t)}{\partial \rho} + 2\lambda q \frac{\partial \mu_{hn}(t)}{\partial \rho}\right)\phi_t.$$

This can be viewed as an ODE in the function $\frac{\partial \phi}{\partial \rho}$ by treating ϕ_t as a fixed function. It can be verified that $0 < \frac{\partial \phi}{\partial \rho} < \infty$ in the limit as $t \to \infty$, that is, in steady state. Furthermore, a simple comparison argument yields that $\partial \mu_{lo}(t)/\partial\rho = \partial\mu_{hn}(t)/\partial\rho < 0$. Hence, the solution to the linear ODE (A.14) is positive since

$$\frac{\partial \phi_t}{\partial \rho} = -\int_t^\infty e^{-R(u-t)}\left(2\lambda(1-q)\frac{\partial \mu_{lo}(u)}{\partial \rho} + 2\lambda q\frac{\partial \mu_{hn}(u)}{\partial \rho}\right)\phi_u\, du > 0.$$

Consider now the case $q = 1$, for which, since $V_{hn} = V_{ln} = 0$,

$$\dot{V}_{ho}(t) = rV_{ho}(t) + \lambda_d \phi_t - 1.$$

Differentiating both sides with respect to ρ and using arguments as above, we see that $\partial V_{ho}(t)/\partial\rho < 0$ since $\partial\phi_t/\partial\rho > 0$. Consequently, $V_{lo}(t) = V_{ho}(t) - \phi_t$ also decreases in ρ.

If $q = 0$, then (A.5) shows that V_{lo} and V_{ho} are independent of ρ. Furthermore,

$$\dot{V}_{ln}(t) = rV_{ln}(t) + \lambda_u(\phi_t - \Delta V_o(t)).$$

As above, we differentiate with respect to ρ and conclude that $V_{ln}(t)$ decreases in ρ since $\partial\phi_t/\partial\rho > 0$ and $\Delta V_o(t)$ is independent of ρ. Consequently, $V_{hn}(t) = V_{ln}(t) - \phi_t + \Delta V_o(t)$ also decreases in ρ. *Q.E.D.*

PROOF OF THEOREM 7:
To see that $z^I > 0$, we note that with $\rho = \rho^C(z)$,

$$\frac{d}{dz}w^I\bigg|_{z=0} = -\delta E\int_0^\infty \frac{d}{d\rho}\mu_{lo}(t,\rho)e^{-rt}\,dt\,\frac{d\rho^C}{dz} > 0,$$

where we have used that $\rho^C(0) = 0$, that $\partial\rho^C/\partial z > 0$ at $z = 0$ (see (A.13)), that $A - B = 0$ if $z = 0$, and that for all t, $\frac{d}{d\rho}\mu_{lo}(t,\rho) < 0$.

To prove that $z^I < z^S \leq z^M = 1$, it suffices to show that the marketmaker welfare is increasing in z, which follows from

$$\begin{aligned}\frac{d}{dz}w^M &= \rho\frac{d}{dz}\left[E\int_0^\infty \mu_{lo}(a-b)e^{-rt}\,dt\right]\\ &= \frac{\rho}{r}\frac{d}{dz}\Gamma'(\rho^C(z))\\ &= \frac{\rho}{r}\Gamma''(\rho^C(z))\frac{d\rho^C}{dz} > 0,\end{aligned}$$

suppressing the arguments t and ρ from the notation, where we have used twice the fact that $\Gamma'(\rho) = rE\int_0^\infty \mu_{lo}(A-B)e^{-rt}\,dt$ if $\rho = \rho^C(z)$ and that $\partial\rho^C/\partial z > 0$ (Theorem 5). *Q.E.D.*

REFERENCES

AMIHUD, Y., AND H. MENDELSON (1980): "Dealership Markets: Market Making with Inventory," *Journal of Financial Economics*, 8, 31–53.

——— (1986): "Asset Pricing and the Bid–Ask Spread," *Journal of Financial Economics*, 17, 223–249.

BAGEHOT, W. (1971): "The Only Game in Town," *The Financial Analysts Journal*, 27, 12–14.

BARCLAY, M. J., W. G. CHRISTIE, J. H. HARRIS, E. KANDEL, AND P. H. SCHULTZ (1999): "Effects of Market Reform on the Trading Costs and Depths of Nasdaq Stocks," *Journal of Finance*, 54, 1–34.

BESTER, H. (1988): "Bargaining, Search Costs and Equilibrium Price Distributions," *Review of Economic Studies*, 55, 201–214.

——— (1989): "Noncooperative Bargaining and Spatial Competition," *Econometrica*, 57, 97–113.

BHATTACHARYA, S., AND K. M. HAGERTY (1987): "Dealerships, Trading Externalities, and General Equilibrium," in *Contractual Arrangements for Intertemporal Trade*, Minnesota Studies in Macroeconomics Series, Vol. 1., ed. by E. Prescott and N. Wallace. Minneapolis: University of Minnesota Press, 81–104.

BILLINGSLEY, P. (1986): *Probability and Measure* (Second Ed.). New York: John Wiley & Sons.

BINMORE, K. G., AND M. J. HERRERO (1988): "Matching and Bargaining in Dynamic Markets," *Review of Economic Studies*, 55, 17–31.

BIRKHOFF, G., AND G.-C. ROTA (1969): *Ordinary Differential Equations*. New York: John Wiley & Sons.

BOARD, S., AND J. ZWIEBEL (2003): "Endogenous Competitive Bargaining," Working Paper, Stanford University.

CONSTANTINIDES, G. M. (1986): "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy*, 94, 842–862.

DAI, Q., AND K. RYDQVIST (2003): "How Do Buyers and Sellers Divide the Surplus," Unpublished Working Paper, Binghamton University.

DEMARZO, P., M. FISHMAN, AND K. HAGERTY (2000): "The Enforcement Policy of a Self-Regulatory Organization," Unpublished Working Paper, Graduate School of Business, Stanford University; forthcoming in the *REStud*.

DIAMOND, P. (1982): "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, 90, 881–894.

DUFFIE, D., N. GÂRLEANU, AND L. H. PEDERSEN (2003): "Valuation in Over-the-Counter Markets," Working Paper, Graduate School of Business, Stanford University.

DUFFIE, D., AND Y. SUN (2004): "The Exact Law of Large Numbers for Pairwise Random Matching," Unpublished Working Paper, Graduate School of Business, Stanford University.

FERLAND, R., AND G. GIROUX (2002): "Une Approche Probabiliste des Marchés Dynamiques, I," Unpublished Working Paper, Université du Québec à Montréal.

GALE, D. (1986a): "Bargaining and Competition Part I: Characterization," *Econometrica*, 54, 785–806.

——— (1986b): "Bargaining and Competition Part II: Existence," *Econometrica*, 54, 807–818.

——— (1987): "Limit Theorems for Markets with Sequential Bargaining," *Journal of Economic Theory*, 43, 20–54.

GARMAN, M. (1976): "Market Microstructure," *Journal of Financial Economics*, 3, 257–275.

GEHRIG, T. (1993): "Intermediation in Search Markets," *Journal of Economics and Management Strategy*, 2, 97–120.

GLOSTEN, L. R., AND P. R. MILGROM (1985): "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71–100.

GREEN, R. C., B. HOLLIFIELD, AND N. SCHURHOFF (2004): "Financial Intermediation and the Costs of Trading in an Opaque Market," Working Paper, Carnegie Mellon University.

HARRIS, L. E., AND M. S. PIWOWAR (2004): "Municipal Bond Liquidity," Working Paper, The Securities and Exchange Commission.

HARRIS, M. (1979): "Expectations and Money in a Dynamic Exchange Model," *Econometrica*, 47, 1403–1419.

HO, T., AND H. R. STOLL (1981): "Optimal Dealer Pricing under Transactions and Return Uncertainty," *Journal of Financial Economics*, 9, 47–73.

HOSIOS, A. J. (1990): "On the Efficiency of Matching and Related Models of Search and Unemployment," *Review of Economic Studies*, 57, 279–298.

HUANG, R. D., AND H. R. STOLL (1996): "Dealer versus Auction Markets: A Paired Comparison of Execution Costs on NASDAQ and the NYSE," *Journal of Financial Economics*, 41, 313–357.

KIYOTAKI, N., AND R. WRIGHT (1993): "A Search-Theoretic Approach to Monetary Economics," *American Economic Review*, 83, 63–77.

KREPS, D. (1990): *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press.

KRISHNAMURTHY, A. (2002): "The Bond/Old-Bond Spread," *Journal of Financial Economics*, 66, 463–506.

KYLE, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 6, 1315–1335.

LAMOUREUX, C. G., AND C. R. SCHNITZLEIN (1997): "When It's Not the Only Game in Town: The Effect of Bilateral Search on the Quality of a Dealer Market," *Journal of Finance*, 52, 683–712.

MCLENNAN, A., AND H. SONNENSCHEIN (1991): "Sequential Bargaining as a Noncooperative Foundation for Walrasian Equilibrium," *Econometrica*, 59, 1395–1424.

MOEN, E. R. (1997): "Competitive Search Equilibrium," *Journal of Political Economy*, 105, 385–411.

MORESI, S. (1991): "Three Essays in Economic Theory," Ph.D. Thesis, MIT.

MORTENSEN, D. T. (1982): "Property Rights and Efficiency in Mating, Racing, and Related Games," *American Economic Review*, 72, 968–979.

NASH, J. (1950): "The Bargaining Problem," *Econometrica*, 18, 155–162.

NEW YORK STOCK EXCHANGE (2001): *Fact Book 2001*.

PAGANO, M. (1989): "Trading Volume and Asset Liquidity," *Quarterly Journal of Economics*, 104, 255–274.

PROTTER, P. (1990): *Stochastic Integration and Differential Equations*. New York: Springer-Verlag.

RUBINSTEIN, A., AND A. WOLINSKY (1985): "Equilibrium in a Market with Sequential Bargaining," *Econometrica*, 53, 1133–1150.

——— (1987): "Middlemen," *Quarterly Journal of Economics*, 102, 581–594.

SCHULTZ, P. (2001): "Corporate Bond Trading Costs: A Peek Behind the Curtain," *Journal of Finance*, 56, 677–698.

SUN, Y. (2000): "The Exact Law of Large Numbers via Fubini Extension and Characterization of Insurable Risks," Working Paper, Singapore National University; forthcoming in *Journal of Economic Theory*.

TREJOS, A., AND R. WRIGHT (1995): "Search, Bargaining, Money, and Prices," *Journal of Political Economy*, 103, 118–140.

VAYANOS, D. (1998): "Transaction Costs and Asset Prices: A Dynamic Equilibrium Model," *Review of Financial Studies*, 11, 1–58.

VAYANOS, D., AND T. WANG (2002): "Search and Endogenous Concentration of Liquidity in Asset Markets," Working Paper, MIT.

WEILL, P.-O. (2002): "The Liquidity Premium in a Dynamic Bargaining Market," Working Paper, Stanford University.

YAVAŞ, A. (1996): "Search and Trading in Intermediated Markets," *Journal of Economics and Management Strategy*, 5, 195–216.