Valuation in Over-the-Counter Markets

January 2, 2006

Abstract

We provide the impact on asset prices of search-and-bargaining frictions in over-the-counter markets. Under certain conditions, illiquidity discounts are higher when counterparties are harder to find, when sellers have less bargaining power, when the fraction of qualified owners is smaller, or when risk aversion, volatility, or hedging demand are larger. Supply shocks cause prices to jump, and then "recover" over time, with a time signature that is exaggerated by search frictions. We discuss a variety of empirical implications.

JEL Classification: G12.

Many assets, such as mortgage-backed securities, corporate bonds, government bonds, emerging-market debt, bank loans, swaps and many other derivatives, private equity, and real estate, are traded in over-the-counter (OTC) markets. Traders in these markets search for counterparties, incurring opportunity or other costs. When counterparties meet, their bilateral relationship is strategic; prices are set through a bargaining process that reflects each investor's alternatives to immediate trade.

We provide a theory of dynamic asset pricing that directly treats search and bargaining in OTC markets. We show how the explicitly calculated equilibrium allocations and prices depend on investors' search abilities, bargaining powers, and risk aversion, and how the time signature of price reactions to supply or demand shocks depends on the speed with which counterparties interact. We discuss a variety of financial applications and testable implications.

Investors in our model contact one another randomly at some mean rate that reflects search ability. When two agents meet, they bargain over the terms of trade based on endogenously determined outside options. Investors are infinitely lived and gains from trade arise from time-varying costs or benefits of holding assets. We show how the equilibrium bargaining powers of the counterparties are determined by search opportunities, using the approach of Rubinstein and Wolinsky (1985).

We first study how search frictions affect asset prices in a steady-state equilibrium in which agents face idiosyncratic risk and with no aggregate risk. We compute steady-state prices both with risk-neutral and risk-averse agents, and show how risk aversion can be approximated in a risk-neutral setting using "holding costs" that capture the utility losses of suboptimal diversification or hedging.

Naturally, search-based market incompleteness is priced by risk-averse agents with time-varying hedging demands. Indeed, under stated conditions, illiquidity discounts are higher if investors can find each other less easily, if buyers have more bargaining power, if the fraction of qualified owners is smaller, if volatility is higher, or if risk aversion is higher. We also indicate situations in which search frictions can lead naturally to an increase in the price of the asset, conveying to it a search-induced scarcity value.

We introduce "aggregate liquidity shocks," that is, shocks that affect the holding costs of many agents simultaneously. We show that, under certain conditions, when an aggregate liquidity shock occurs, the price drops and recovers only slowly. The speed of the price recovery depends on the search

1

intensity that determines the speed of reallocation of securities to the more liquid agents and on the time that it takes for illiquid agents to become liquid, for example to "recapitalize." Also, the risk of future aggregate liquidity shocks significantly lowers the post-recovery price level. Search frictions thus affect both the general level of prices as well as the resiliency of the market to aggregate shocks. More illiquid markets (those with lower search intensities) have generally lower price levels, larger price reactions to supply shocks, and slower price recovery.

When an aggregate liquidity shock occurs, the expected utilities of asset owners, even those owners who are not directly affected by the liquidity shock themselves, decrease because selling opportunities worsen: Sellers' search times increase and their bargaining positions deteriorate. Conversely, the expected utilities of agents "waiting on the sideline," those with no asset position, increase at times of aggregate liquidity shock, because they may have the opportunity to purchase securities at distressed prices.

We discuss how our results contribute to an explanation of the time signatures of price responses to several types of aggregate liquidity shocks, to corporate bonds that are downgraded or in default (Hradsky and Long (1989)), to sovereign bonds during "debt crises," to individual stocks at index inclusion or exclusion events (as in Greenwood (2005)), to stocks affected by sudden large outside orders (Andrade, Chang, and Seasholes (2005) and Coval and Stafford (2005)), or to catastrophe reinsurance risk premia after large unexpected losses in capital caused by events such as major hurricanes (Froot and O'Connell (1999)), among other relevant empirical phenomena.

The point of departure of this paper is a variant of the basic risk-neutral search-based pricing model of Duffie, Gârleanu, and Pedersen (2005). While Duffie, Gârleanu, and Pedersen (2005) focus on the steady-state pricing of a simple consol bond and treat the behavior of marketmakers and the implications of search frictions for bid-ask spreads, this paper instead treats the implications of search frictions for risky asset pricing. We provide *(i)* the impact on asset prices of risk aversion in a setting with search, above and beyond the usual implications of risk sharing in incomplete markets, *(ii)* the implications of search frictions for the time dynamics of price responses to supply or demand shocks, and *(iii)* the determination of endogenous bargaining power[1] based on the alternative search opportunities of the buyers and

---

[1]Duffie, Gârleanu, and Pedersen (2005) take the bargaining powers of buyers and sellers as exogenous to the model.

sellers.

Search models have been studied extensively in the context of labor economics, starting with the "coconuts" model of Diamond (1982), and in the context of monetary economics, for example, Trejos and Wright (1995). As for search-based models of asset pricing, Weill (2002) and Vayanos and Wang (2002) have extended the risk-neutral version of our model in order to treat multiple assets in the same economy, obtaining cross-sectional restrictions on asset returns. Duffie, Gârleanu, and Pedersen (2005) treat marketmakers, showing that search frictions have different implications for bid-ask spreads than do information frictions. Miao (2004) provides a variant of this model. Weill (2003) studies the implications of search frictions in an extension of our model in which marketmakers' inventories "lean against" the outside order flow. Newman and Rierson (2003) present a model in which supply shocks temporarily depress prices across correlated assets, as providers of liquidity search for long-term investors, supported by empirical evidence of issuance impacts across the European telecommunications bond market. Duffie, Gârleanu, and Pedersen (2002) use a search-based model of the impact on asset prices and securities lending fees of the common institution by which would-be shortsellers must locate lenders of securities before being able to sell short. Difficulties in locating lenders of shares can allow for dramatic price imperfections, as, for example, in the case of the spinoff of Palm, Incorporated, documented by Mitchell, Pulvino, and Stafford (2002) and Lamont and Thaler (2003).

Our results also complement the literature treating the effect on asset prices of an exogenously specified trading cost (Amihud and Mendelson (1986), Constantinides (1986), Vayanos (1998), Huang (2003), and Acharya and Pedersen (2005)) by endogenizing the trading cost in the context of OTC markets. Krainer and LeRoy (2002) study housing prices in a different search framework. Longstaff (2004) addresses market frictions with the device of deterministic "blackout" periods on individual trade.

The remainder of the paper is organized as follows. Section 1 lays out a baseline model with risk-neutral agents. Section 2 then treats an OTC market for a risky asset whose risk-averse owners search for potential buyers when the asset ceases to be a relatively good endowment hedge. We characterize how search frictions magnify risk premia beyond those of a liquid but incomplete-markets setting. Section 3 provides the implications of search frictions for price reactions to supply or demand shocks, showing especially how the time pattern of "price recovery" is influenced by search frictions. Finally, Section 4

describes the empirical implications of search frictions for asset pricing in a range of actual OTC markets. Some proofs and supplementary results are relegated to appendices.

1 Basic Search Model of Asset Prices

This section introduces a baseline risk-neutral model of an over-the-counter market, that is, a market in which agents can trade only when they meet each other, and in which transaction prices are bargained. This baseline model, simplified from Duffie, Gârleanu, and Pedersen (2005) by stripping out marketmakers, is then generalized in the remainder of the paper to treat risk aversion, and the effects of aggregate liquidity shocks.

Agents are risk-neutral and infinitely lived, with a constant time-preference rate $\beta > 0$ for consumption of a single non-storable numeraire good.[2]

An agent can invest in a bank account — which can also be interpreted as a "liquid" security — with a risk-free interest rate of r. As a form of credit constraint that rules out "Ponzi schemes," the agent must enforce some lower bound on the liquid wealth process W. We take $r = \beta$ in this baseline model, since agents are risk neutral.

Agents may trade a long-lived asset in an over-the-counter market. The asset can be traded only bilaterally, when in contact with a counterparty. We begin for simplicity by taking the OTC asset to be a consol, which pays one unit of consumption per unit of time. Later, when introducing the effect of risk aversion, we generalize to random dividend processes.

An agent is characterized by an intrinsic preference for asset ownership that is "high" or "low." A low-type agent, when owning the asset, has a holding cost of δ per time unit. A high-type agent has no such holding cost. We could imagine this holding cost to be a shadow price for ownership due, for example, to *(i)* low liquidity, that is, a need for cash, *(ii)* high financing or financial-distress costs, *(iii)* adverse correlation of asset returns with endowments (formalized in Section 2), *(iv)* a relative tax disadvantage, as studied by Dai and Rydqvist (2003) in an empirical analysis of search-and-bargaining effects in the context of tax trading,[3] or *(v)* a relatively low

---

[2]Specifically, an agent's preferences among adapted finite-variation cumulative consumption processes are represented by the utility $E\left(\int_0^\infty e^{-\beta t}\, dC_t\right)$ for a cumulative consumption process C, whenever the integral is well defined.

[3]Dai and Rydqvist (2003) study tax trading between a small group of foreign investors

personal use for the asset, as may happen, for example, for certain durable consumption goods such as homes. The agent's intrinsic type is a Markov chain, switching from low to high with intensity λ_u, and back with intensity λ_d. The intrinsic-type processes of any two agents are independent.[4]

A fraction s of agents are initially endowed with one unit of the asset. Investors can hold at most one unit of the asset and cannot shortsell. Because agents have linear utility, it is without much loss of generality that we restrict attention to equilibria in which, at any given time and state of the world, an agent holds either 0 or 1 unit of the asset. Hence, the full set of agent types is $\mathcal{T} = \{ho, hn, lo, ln\}$, with the letters "$h$" and "$l$" designating the agent's current intrinsic liquidity state as high or low, respectively, and with "o" or "n" indicating whether the agent currently owns the asset or not, respectively.

We suppose that there is a "continuum" (a non-atomic finite measure space) of agents, and let $\mu_\sigma(t)$ denote the fraction at time t of agents of type $\sigma \in \mathcal{T}$, so that

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

Equating the per-capita supply s with the fraction of owners gives

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

An agent finds a counterparty with an intensity λ, reflecting the efficiency of the search technology. We assume the counterparty found is randomly selected from the pool of other agents, so that the probability that the counterparty is of type σ is $\mu_\sigma(t)$. Thus, the total intensity of finding a type-σ investor is $\lambda\mu_\sigma$. Hence, assuming that the law of large numbers applies, hn investors contact lo investors at a total (almost sure) rate of $\lambda\mu_{lo}\mu_{hn}$ and, since lo investors contact hn investors at the same total rate, the total rate of such counterparty matchings is $2\lambda\mu_{lo}\mu_{hn}$. Duffie and Sun (2004) provide

---

and a larger group of domestic investors. They find that investors from the "long side of the market" get part of the gains from trade, under certain conditions, which they interpret as evidence of a search-and-bargaining equilibrium.

[4]All random variables are defined on a probability space $(\Omega, \mathcal{F}, Pr)$ with corresponding filtration $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-algebras of $\mathcal{F}$ satisfying the usual conditions, as defined by Protter (1990). The filtration represents the resolution over time of information commonly available to investors.

a discrete-time search-and-matching model in which the exact law of large numbers for a continuum of agents indeed applies in this sense.[5]

To solve the model, we proceed in two steps. First, we use the insight that the only form of encounter that provides gains from trade is one in which low-type owners sell to high-type non-owners. From bargaining theory, we know (see Appendix A) that at these encounters, trade occurs immediately. We can therefore determine the asset allocations without reference to prices. Given the time-dynamics of the masses, $\{\mu(t) : t \geq 0\}$, we then consider an investor's lifetime utility, depending on the investor's type, the bargaining problem, and the resulting price.

In equilibrium, the rates of change of the fractions of the respective investor types are

$$
\begin{aligned}
\dot{\mu}_{lo}(t) &= -2\lambda\mu_{hn}(t)\mu_{lo}(t) - \lambda_u\mu_{lo}(t) + \lambda_d\mu_{ho}(t) \\
\dot{\mu}_{hn}(t) &= -2\lambda\mu_{hn}(t)\mu_{lo}(t) - \lambda_d\mu_{hn}(t) + \lambda_u\mu_{ln}(t) \\
\dot{\mu}_{ho}(t) &= 2\lambda\mu_{hn}(t)\mu_{lo}(t) - \lambda_d\mu_{ho}(t) + \lambda_u\mu_{lo}(t) \\
\dot{\mu}_{ln}(t) &= 2\lambda\mu_{hn}(t)\mu_{lo}(t) - \lambda_u\mu_{ln}(t) + \lambda_d\mu_{hn}(t).
\end{aligned}
\tag{3}
$$

The intuition for, say, the first equation in (3) is straightforward: Whenever an lo agent meets an hn investor, he sells his asset and is no longer an lo agent. This (together with the law of large numbers) explains the first term on the right hand side of (3). The second term is due to intrinsic type changes in which lo investors become ho investors, and the last term is due to intrinsic type changes from ho to lo.

Duffie, Gârleanu, and Pedersen (2005) show that there is a unique stable steady-state solution for $\{\mu(t) : t \geq 0\}$, that is, a constant solution defined by $\dot{\mu}(t) = 0$. The steady state is computed by using (1)–(2) and the fact that $\mu_{lo} + \mu_{ln} = \lambda_d/(\lambda_u + \lambda_d)$, in order to write the first equation in (3) as a quadratic equation in μ_{lo}, given as Appendix equation (C.1).

Having determined the steady-state fractions of investor types, we compute the investors' equilibrium intensities of finding counterparties of each type and, hence, their utilities for remaining lifetime consumption, as well as the bargained price P. The utility of a particular agent depends on his current type, $\sigma(t) \in \mathcal{T}$, and the wealth $W(t)$ in his bank account. Specifically,

---

[5]Giroux (2005) proves that the cross-sectional distribution of agent types in a natural discrete-time analogue of this model indeed converges to the continuous-time model studied here.

6

lifetime utility is $W(t) + V_{\sigma(t)}$, where, for each investor type σ in $\mathcal{T}$, V_σ is a constant to be determined.

In steady state, the rate of growth of any agent's expected indirect utility must be the discount rate r, which yields the steady-state equations

$$
\begin{aligned}
0 &= rV_{lo} - \lambda_u(V_{ho} - V_{lo}) - 2\lambda\mu_{hn}(P - V_{lo} + V_{ln}) - (1 - \delta) \\
0 &= rV_{ln} - \lambda_u(V_{hn} - V_{ln}) \\
0 &= rV_{ho} + \lambda_d(V_{ho} - V_{lo}) - 1 \\
0 &= rV_{hn} + \lambda_d(V_{hn} - V_{ln}) - 2\lambda\mu_{lo}(V_{ho} - V_{hn} - P).
\end{aligned}
\tag{4}
$$

The price is determined through bilateral bargaining. A high-type non-owner pays at most his reservation value $\Delta V_h = V_{ho} - V_{hn}$ for obtaining the asset, while a low-type owner requires a price of at least $\Delta V_l = V_{lo} - V_{ln}$. Nash bargaining, or the Rubinstein-type game considered in Appendix A, implies that the bargaining process results in the price

$$
P = \Delta V_l(1 - q) + \Delta V_h\, q\,,
\tag{5}
$$

where $q \in [0, 1]$ is the bargaining power of the seller.

While Nash equilibrium is consistent with exogenously assumed bargaining powers, Appendix A applies the device of Rubinstein and Wolinsky (1985) to calculate the unique bargaining powers that represent the limiting prices of a sequence of economies in which, once a pair of counteparties meets to negotiate, one of the pair is selected at random to make an offer to the other, at each of a sequence of offer times separated by intervals that shrink to zero. Specifically, suppose that when two agents find each other, one of them is chosen randomly, the seller with probability $\hat{q}$ and the buyer with probability $1 - \hat{q}$, to suggest a trading price. The other either rejects or accepts the offer, immediately. If the offer is rejected, the owner receives the dividend from the asset during the current period. At the next period, Δ_t later, one of the two agents is chosen at random, independently, to make a new offer. The bargaining may, however, break down before a counteroffer is made. A breakdown may occur because at least one of the agents changes valuation type, or if one of the agents meets yet another agent, and leaves his current trading partner, provided agents can indeed continue to search while engaged in negotiation. In that case, as shown in Appendix A, the limiting price as Δ_t goes to zero is represented by (5), with the bargaining power of the seller q equal to $\hat{q}$. This simple solution, in which the only "bargaining advantage" that matters in the limit is the likelihood of being selected as the agent that

makes the next offer, arises because a counterparty's ability to meet an alternative trading partner while negotiating makes that counterparty more impatient, but also increases the trading partner's risk of breakdown, to the point that these two effects are precisely offseting.

If, however, agents cannot search for alternative trading partners during negotiations, then the limiting price is that associated with the bargaining power

$$q = \frac{\hat{q}(r + \lambda_u + \lambda_d + 2\lambda\mu_{lo})}{\hat{q}(r + \lambda_u + \lambda_d + 2\lambda\mu_{lo}) + (1 - \hat{q})(r + \lambda_u + \lambda_d + 2\lambda\mu_{hn})}. \tag{6}$$

For the comparative statics that follow, we will use the limiting bargaining power associated with search during negotiation, in order to simplify the analysis by avoiding the dependence in (6) of the seller's bargaining power q on various parameters that may be shifted as part of the experiment being considered.

The linear system (4)-(5) of equations has a unique solution, with

$$P = \frac{1}{r} - \frac{\delta}{r}\frac{r(1 - q) + \lambda_d + 2\lambda\mu_{lo}(1 - q)}{r + \lambda_d + 2\lambda\mu_{lo}(1 - q) + \lambda_u + 2\lambda\mu_{hn}q}. \tag{7}$$

This price (7) is the present value, $1/r$, of dividends, reduced by an illiquidity discount. The price is lower and the discount is larger, *ceteris paribus*, if the distressed owner has less hope of switching type (lower λ_u), if it is more difficult for the owner to find other buyers (lower μ_{hn}), if the buyer may more suddenly need liquidity himself (higher λ_d), if it is easier for the buyer to find other sellers (higher μ_{lo}), or if the seller has less bargaining power (lower q).

These intuitive results are based on partial derivatives of the right-hand side of (7) — in other words, they hold when a parameter changes without influencing any of the others. We note, however, that the steady-state type fractions μ themselves depend on λ_d, λ_u, and λ. The following proposition offers a characterization of the equilibrium steady-state effect of changing each parameter.

Proposition 1 *The steady-state equilibrium price P is decreasing in δ, s, and λ_d, and is increasing in λ_u and q. Further, if $s < \lambda_u/(\lambda_u + \lambda_d)$, then $P \to 1/r$ as $\lambda \to \infty$, and P is increasing in λ for all $\lambda \geq \bar{\lambda}$, for a constant $\bar{\lambda}$ depending on the other parameters of the model.*

The condition that $s < \lambda_u/(\lambda_u + \lambda_d)$ means that, in steady state, there is less than one unit of asset per agent of high intrinsic type. While this corresponds to the intuitively anticipated increase in market value with increasing bilateral contact rate, the alternative is also possible. With $s > \lambda_u/(\lambda_u + \lambda_d)$, the marginal investor in perfect markets has the relatively lower reservation value, and search frictions lead to a "scarcity value." For example, a high-type investor in an illiquid OTC market could pay more than the Walrasian price for the asset because it is hard to find, and given no opportunity to exploit the effect of immediate competition among many sellers. This scarcity value could, for example, contribute to the widely studied on-the-run premium for Treasuries, as discussed in Section 4.

It can be checked that the above results extend to risky dividends in at least the following senses: *(i)* If the cumulative dividend is risky with constant drift ν, then the equilibrium is that for a consol bond with dividend rate of ν; *(ii)* if the dividend rate and illiquidity cost are proportional to a process X with $E_t[X(t+u)] = X(t)e^{\nu u}$, for some constant growth rate ν, then the price and value functions are also proportional to X, with factors of proportionality given as above, with r replaced by $r - \nu$; *(iii)* if the dividend-rate process X satisfies $E_t[X(t+u)] = X(t) + mu$ for a constant drift m (and if illiquidity costs are constant), then the continuation values are of the form $X(t)/r + v_\sigma$ for owners and v_σ for non-owners, where the constants v_σ are computed in a similar manner.

Next, we model risky dividends, using case *(i)* above, in the context of risk-averse agents. In a previous version of the paper, we described how one can use case *(iii)* in a model in which investors have risk limits based on value at risk.

2 Risk-Aversion

This section provides a version of the asset-pricing model with risk aversion, in which the motive for trade between two agents is the different extent to which they derive hedging benefits from owning the asset. We provide a sense in which this economy can be interpreted in terms of the baseline economy of Section 1.

Agents have constant-absolute-risk-averse (CARA) additive utility, with a coefficient γ of absolute risk aversion and with time preference at rate β.

An asset has a cumulative dividend process D satisfying

$$dD(t) = m_D \, dt + \sigma_D \, dB(t), \tag{8}$$

where m_D and σ_D are constants, and B is a standard Brownian motion with respect to the given probability space and filtration $(\mathcal{F}_t)$. Agent i has a cumulative endowment process η^i, with

$$d\eta^i(t) = m_\eta \, dt + \sigma_\eta \, dB^i(t), \tag{9}$$

where the standard Brownian motion B^i is defined by

$$dB^i(t) = \rho^i(t) \, dB(t) + \sqrt{1 - \rho^i(t)^2} \, dZ^i(t), \tag{10}$$

for a standard Brownian motion Z^i independent of B, and where $\rho^i(t)$ is the "instantaneous correlation" between the asset dividend and the endowment of agent i. We model ρ^i as a two-state Markov chain with states ρ_h and $\rho_l > \rho_h$. The intrinsic type of an agent is identified with this correlation parameter. An agent i whose intrinsic type is currently high (that is, with $\rho^i(t) = \rho_h$) values the asset more highly than does a low-intrinsic-type agent, because the increments of the high-type endowment have lower conditional correlation with the asset's dividends. As in the baseline model of Section 1, agents' intrinsic types are pairwise-independent Markov chains, switching from l to h with intensity λ_u, and from h to l with intensity λ_d. An agent owns either θ_n or θ_o units of the asset, where $\theta_n < \theta_o$. For simplicity, no other positions are permitted, which entails a loss in generality. Agents can trade the OTC security only when they meet, as previously. The agent type space is $\mathcal{T} = \{lo, ln, ho, hn\}$. In this case, the symbols 'o' and 'n' indicate large and small owners, respectively. Given a total supply Θ of shares per investor, market clearing requires that

$$(\mu_{lo} + \mu_{ho})\theta_o + (\mu_{ln} + \mu_{hn})\theta_n = \Theta, \tag{11}$$

which, using (1), implies that the fraction of large owners is

$$\mu_{lo} + \mu_{ho} = s \equiv \frac{\Theta - \theta_n}{\theta_o - \theta_n}. \tag{12}$$

We consider a particular agent whose type process is $\{\sigma(t) : t \geq 0\}$, and let θ denote the associated asset-position process (that is, $\theta(t) = \theta_o$ whenever

$\sigma(t) \in \{ho, lo\}$ and otherwise $\theta(t) = \theta_n$). We suppose that there is a perfectly liquid "money-market" asset with a constant risk-free rate of return r, which, for simplicity, is assumed to be determined outside of the model, and with a perfectly elastic supply, as is typical in the literature treating multi-period asset-pricing models based on CARA utility, such as Wang (1994). The agent's money-market wealth process W therefore satisfies

$$dW(t) = (rW(t) - c(t))\, dt + \theta(t)\, dD(t) + d\eta(t) - P\, d\theta(t),$$

where c is the agent's consumption process, η is the agent's cumulative endowment process, and P is the asset price per share (which is constant in the equilibria that we examine). The last term thus captures payments in connection with trade. The consumption process is required to satisfy measurability, integrability, and transversality conditions stated in Appendix C.

We consider a steady-state equilibrium, and let $J(w, \sigma)$ denote the indirect utility of an agent of type $\sigma \in \{lo, ln, ho, hn\}$ with current wealth w. Assuming sufficient differentiability, the Hamilton-Jacobi-Bellman (HJB) equation for an agent of current type lo is

$$
\begin{aligned}
0 = \sup_{\overline{c} \in \mathbb{R}} \{ &-e^{-\gamma \overline{c}} + J_w(w, lo)(rw - \overline{c} + \theta_o m_D + m_\eta) \\
&+ \frac{1}{2} J_{ww}(w, lo)(\theta_o^2 \sigma_D^2 + \sigma_\eta^2 + 2\rho_l \theta_o \sigma_D \sigma_\eta) - \beta J(w, lo) \\
&+ \lambda_u[J(w, ho) - J(w, lo)] + 2\lambda \mu_{hn}[J(w + P\overline{\theta}, ln) - J(w, lo)] \},
\end{aligned}
$$

where $\overline{\theta} = \theta_o - \theta_n$. The HJB equations for the other agent types are similar. Under technical regularity conditions found in Appendix C, we verify that

$$J(w, \sigma) = -e^{-r\gamma(w + a_\sigma + \overline{a})}, \tag{13}$$

where

$$\overline{a} = \frac{1}{r}\left(\frac{\log r}{\gamma} + m_\eta - \frac{1}{2}r\gamma\sigma_\eta^2 - \frac{r - \beta}{r\gamma}\right), \tag{14}$$

and where, for each σ, the constant a_σ is determined as follows. The first-order conditions of the HJB equation of an agent of type σ imply an optimal consumption rate of

$$\overline{c} = -\frac{\log(r)}{\gamma} + r(w + a_\sigma + \overline{a}). \tag{15}$$

Inserting this solution for $\bar{c}$ into the respective HJB equations yields a system of equations characterizing the coefficients a_σ.

The price P is determined using Nash bargaining with seller bargaining power q, similar in spirit to the baseline model of Section 1. Given the reservation values of buyer and seller implied by $J(w, \sigma)$, the bargaining price satisfies $a_{lo} - a_{ln} \leq P\bar{\theta} \leq a_{ho} - a_{hn}$. The following result obtains.

Proposition 2 *In equilibrium, an agent's consumption rate is given by (15), the value function is given by (13), and $(a_{lo}, a_{ln}, a_{ho}, a_{hn}, P) \in \mathbb{R}^5$ solve*

$$0 = ra_{lo} + \lambda_u \frac{e^{-r\gamma(a_{ho}-a_{lo})} - 1}{r\gamma} + 2\lambda\mu_{hn} \frac{e^{-r\gamma(P\bar{\theta}+a_{ln}-a_{lo})} - 1}{r\gamma} - (\kappa(\theta_o) - \theta_o\bar{\delta})$$

$$0 = ra_{ln} + \lambda_u \frac{e^{-r\gamma(a_{hn}-a_{ln})} - 1}{r\gamma} - (\kappa(\theta_n) - \theta_n\bar{\delta}) \tag{16}$$

$$0 = ra_{ho} + \lambda_d \frac{e^{-r\gamma(a_{lo}-a_{ho})} - 1}{r\gamma} - \kappa(\theta_o)$$

$$0 = ra_{hn} + \lambda_d \frac{e^{-r\gamma(a_{ln}-a_{hn})} - 1}{r\gamma} + 2\lambda\mu_{lo} \frac{e^{-r\gamma(-P\bar{\theta}+a_{ho}-a_{hn})} - 1}{r\gamma} - \kappa(\theta_n),$$

with

$$\kappa(\theta) = \theta m_D - \frac{1}{2}r\gamma\left(\theta^2\sigma_D^2 + 2\rho_h\theta\sigma_D\sigma_\eta\right) \tag{17}$$

$$\bar{\delta} = r\gamma(\rho_l - \rho_h)\sigma_D\sigma_\eta > 0, \tag{18}$$

as well as the Nash bargaining equation,

$$q\left(1 - e^{r\gamma(P\bar{\theta}-(a_{lo}-a_{ln}))}\right) = (1 - q)\left(1 - e^{r\gamma(-P\bar{\theta}+a_{ho}-a_{hn})}\right). \tag{19}$$

A natural benchmark is the limit price associated with vanishing search frictions, characterized as follows.

Proposition 3 *If $s < \mu_{hn} + \mu_{ho}$, then, as $\lambda \to \infty$,*

$$P \to \frac{\kappa(\theta_o) - \kappa(\theta_n)}{r\bar{\theta}}. \tag{20}$$

In order to compare the equilibrium for this model to that of the baseline model, we use the linearization $e^z - 1 \approx z$, which leads to

$$
\begin{aligned}
0 &\approx ra_{lo} - \lambda_u(a_{ho} - a_{lo}) - 2\lambda\mu_{hn}(P\bar{\theta} - a_{lo} + a_{ln}) - (\kappa(\theta_o) - \theta_o\bar{\delta}) \\
0 &\approx ra_{ln} - \lambda_u(a_{hn} - a_{ln}) - (\kappa(\theta_n) - \theta_n\bar{\delta}) \\
0 &\approx ra_{ho} - \lambda_d(a_{lo} - a_{ho}) - \kappa(\theta_o) \\
0 &\approx ra_{hn} - \lambda_d(a_{ln} - a_{hn}) - 2\lambda\mu_{lo}(a_{ho} - a_{hn} - P\bar{\theta}) - \kappa(\theta_n) \\
P\bar{\theta} &\approx (1 - q)(a_{lo} - a_{ln}) + q(a_{ho} - a_{hn}).
\end{aligned}
$$

These equations are of the same form as those in Section 1 for the indirect utilities and asset price in an economy with risk-neutral agents, with dividends at rate $\kappa(\theta_o)$ for large owners and dividends at rate $\kappa(\theta_n)$ for small owners, and with illiquidity costs given by $\bar{\delta}$ of (18). In this sense, we can view the baseline model as a risk-neutral approximation of the effect of search illiquidity in a model with risk aversion. The approximation error goes to zero for small agent heterogeneity (that is, small $\rho_l - \rho_h$).[6] Solving specifically for the price P in the associated linear model, we have

$$
P = \frac{\kappa(\theta_o) - \kappa(\theta_n)}{r\bar{\theta}} - \frac{\bar{\delta}}{r}\frac{r(1 - q) + \lambda_d + 2\lambda\mu_{lo}(1 - q)}{r + \lambda_d + 2\lambda\mu_{lo}(1 - q) + \lambda_u + 2\lambda\mu_{hn}q}. \quad (21)
$$

The price is the sum of the perfect-liquidity price (that for the case of $\lambda = +\infty$), plus an adjustment for illiquidity that can be viewed as the present value of a perpetual stream of risk premia that are due to search frictions. The illiquidity component depends on the strength of the difference in hedging motives for trade by two types of agents, in evidence in the factor $\bar{\delta}$ defined by (18). One of these types of agents can be viewed as the natural hedger; the other can be viewed as the type that provides the hedge, at an extra risk premium. The illiquidity risk premium need not be increasing in the degree of overall "market risk" exposure of the asset, and would be non-zero even if there were no aggregate endowment risk.[7] Graveline and McBrady

---

[6]The error introduced by the linearization is in $O\big((a_{ho} - a_{lo})^2 + (a_{hn} - a_{ln})^2 + (P\bar{\theta} - a_{lo} + a_{ln})^2\big)$, which, by continuity, is in $O\big((\rho_l - \rho_h)^2\big)$ for a compact parameter space. Hence, if $\rho_l - \rho_h$ is small, then the approximation error is an order of magnitude smaller, of the order $(\rho_l - \rho_h)^2$.

[7]We could arrange for the absence of aggregate endowment risk, for example by having half the population exposed positively to the asset, the other half exposed negatively, in an offsetting way, with the portions of endowment risks that are orthogonal to the asset returns being idiosyncratic and adding up (by the law of large numbers) to zero. (We

(2005) empirically link the size of repo specials in on-the-run treasuries to the motives of financial services firms to hedge their inventories of corporate and mortgage-backed securities. The repo specials, which are reflections of search frictions in the treasury repo market, are shown to be larger when the inventories are larger, and larger when interest-rate volatility is higher, consistent with (18).

Numerical Example. We select parameters for a numerical illustration of the implications of the model for a market with an annual asset turnover rate of about 50%, which is roughly that of the over-the-counter market for corporate bonds. Table 1 contains the exogenous parameters for the base-case risk-neutral model, and Table 2 contains the resulting steady-state fractions of each type and the price. The search intensity of $\lambda = 625$ shown in Table 1 implies that an agent expects to be in contact with $2\lambda = 1250$ other agents each year, that is, $1250/250 = 5$ agents a day. Given the equilibrium mass of potential buyers, the average time needed to sell is $250 \times (2\lambda\mu_{hn})^{-1} = 1.8$ days. The switching intensities λ_u and λ_d mean that a high-type investor remains a high type for an average of 2 years, while an illiquid low type remains a low type for an average of 0.2 years. These intensities imply an annual turnover of $2\lambda\mu_{lo}\mu_{hn}/s = 49\%$ which roughly matches the median annual bond turnover of 51.7% reported by Edwards, Harris, and Piwowar (2004). The fraction of investors holding a position is $s = 0.8$, the discount and interest rates are 5%, sellers and buyers each have half of the bargaining power $q = 0.5$, and the illiquidity cost is $\delta = 2.5$, as implied by the risk aversion parameters discussed below.

We see that only a small fraction of the asset, $\mu_{lo}/s = 0.0028/0.8 = 0.35\%$ of the total supply, is mis-allocated to low intrinsic types because of search frictions. The equilibrium asset price, 18.38, however, is substantially below the perfect market price of $1/r = 20$, reflecting a significant impact of illiquidity on the price, despite the relatively small impact on the asset allocation. Stated differently, we can treat the asset as a bond whose yield, dividend rate of 1 divided by price, is $1/18.38 = 5.44\%$, or 44 basis points above the liquid-market yield r. This yield spread is of the order of magnitude of the corporate-bond liquidity spread estimated by Longstaff, Mithal, and

can adjust our model so that the asset is held in zero net supply, allowing short and long positions; this was done in the risk-limits section.)

λ	λ_u	λ_d	s	r	β	q	δ
625	5	0.5	0.80	0.05	0.05	0.5	2.5

Table 1: Base-case parameters for baseline model.

μ_{ho}	μ_{hn}	μ_{lo}	μ_{ln}	P
0.7972	0.1118	0.0028	0.0882	18.38

Table 2: Steady-state masses and asset price, baseline model.

Neis (2004), of between 9 and 65 basis points, depending on the specification and reference risk-free rate.

The base-case risk-neutral model specified in Table 1 corresponds to a model with risk-averse agents with additional parameters given in Table 3 in the following sense. First, the "illiquidity cost" $\delta = \overline{\delta} = 2.5$ of low-intrinsic-type is that implied by (18) from the hedging costs of the risk-aversion model. Second, the total amount Θ of shares and the investor positions, θ_o and θ_n, imply the same fraction $s = 0.8$ of the population holding large positions, using (12). The investor positions that we adopt for this calibration are realistic in light of the positions adopted by high and low type investors in the associated Walrasian (perfect) market with unconstrained trade sizes, which, as shown in Appendix B, has an equilibrium large-owner position size of 17,818 shares and a small-owner position size of $-2,182$ shares. Third, the certainty-equivalent dividend-rate per share, $(\kappa(\theta_o) - \kappa(\theta_n))/(\theta_o - \theta_n) = 1$, is the same as that of the baseline model. Finally, the mean parameter $\mu_D = 1$ and volatility parameter $\sigma_D = 0.5$ of the asset's risky dividend implies that the standard deviation of yearly returns on the bond is approximately $\sigma_D/P = 2.75\%$.

γ	ρ_h	ρ_l	μ_η	σ_η	μ_D	σ_D	Θ	θ_o	θ_n
0.01	-0.5	0.5	10000	10000	1	0.5	16000	20000	0

Table 3: Additional base-case parameters with risk-aversion.

Figure 1 shows how prices increase with liquidity, as measured by the search intensity λ. The graph reflects the fact that, as the search intensity
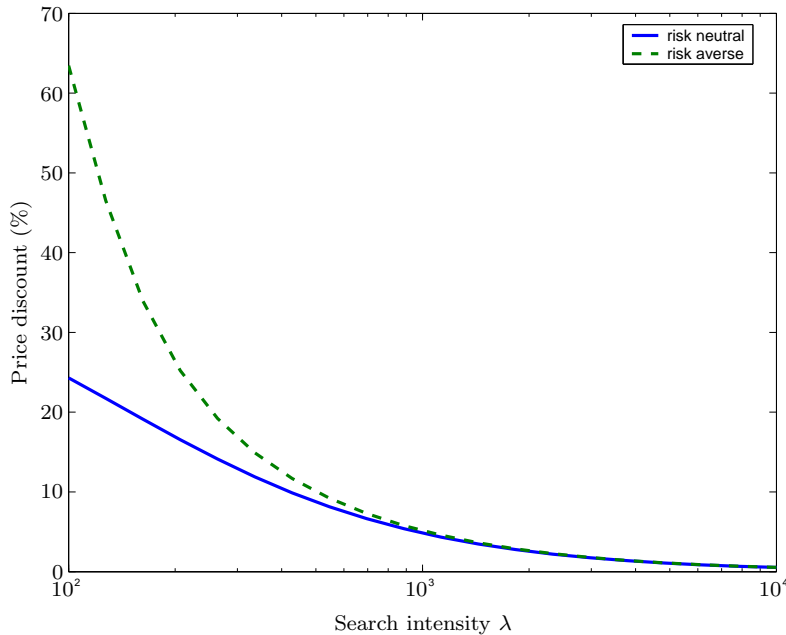
Figure 1: Proportional price reduction relative to the perfect-market price, as a function of the search intensity λ.

λ becomes large, the allocation and price converge to their perfect-market counterparts (Propositions 1 and 3).

Figures 2 and 3 show how prices are discounted for illiquidity, relative to the perfect-markets price, by an amount that depends on risk aversion and volatility. As we vary the parameters in these figures, we compute both the equilibrium solution of the risk-aversion model and the solution of the associated baseline risk-neutral model that is obtained by the linearization (21), taking $\overline{\delta}$ from (18) case by case.

We see that the illiquidity discount increases with risk aversion and volatility, and that both effects are large for our benchmark parameters. The illiquidity discount ranges between 1% and 40%, depending on the risk and risk aversion.

These figures also show that the equilibrium price of the OTC market model with risk aversion is generally well approximated by our closed-form expression (21).
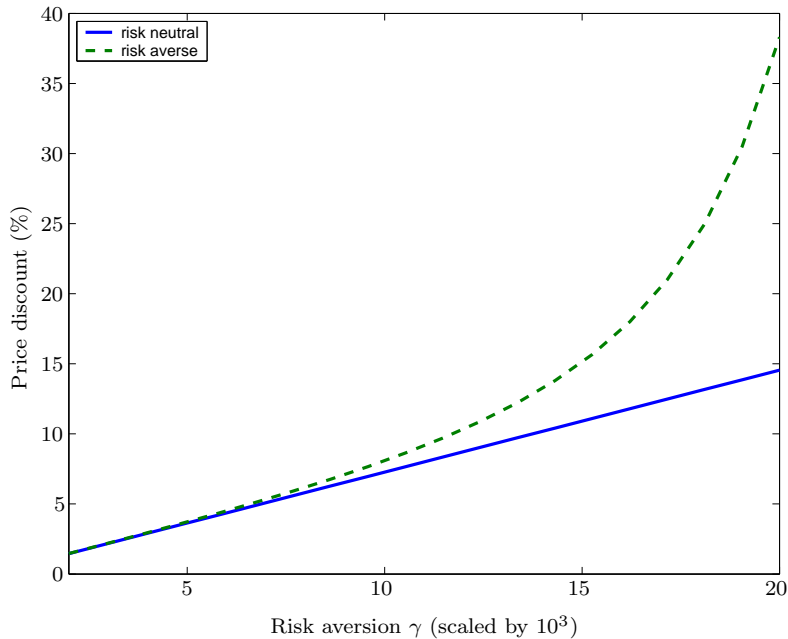
16

Figure 2: Proportional price reduction relative to the perfect-market price, as a function of the investor risk aversion γ. The dashed line corresponds to the model with risk-averse agents (Equations (16)–(19)). The solid line corresponds to the linearized model (Equation (21)), in which the parameters $\bar{\delta}$ and κ change with γ.

3   Aggregate Liquidity Shocks

So far, we have studied how search frictions affect steady-state prices and returns in a setting in which agents receive idiosyncratic liquidity shocks, with no macro-uncertainty.

Search frictions affect not only the average levels of asset prices, but also the asset market's resilience to aggregate shocks. We examine this by characterizing the impact of aggregate liquidity shocks that simultaneously affect many agents. We are interested in the shock's immediate effect on prices, the time-pattern of the price recovery, the ex-ante price effect due to the risk of future shocks, and the change in equilibrium search times.

Our results are broadly consistent with the description by Froot and O'Connell (1999) of the behavior of catastrophe reinsurance risk premia after large unexpected losses in capital caused by events such as major hurricanes. After large initial jumps in reinsurance premia caused by the sudden
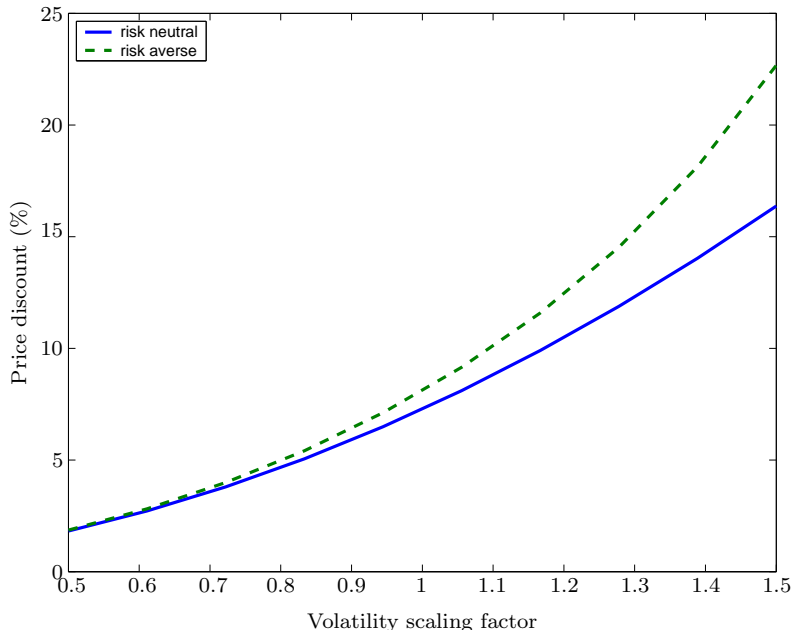
Figure 3: Proportional price reduction relative to the perfect-market price, as a function of a volatility scaling factor that scales both σ_η and σ_D. The dashed line corresponds to the model with risk-averse agents (Equations (16)–(19)). The solid line corresponds to the linearized model (Equation (21)), in which the parameters $\bar{\delta}$ and κ change with σ_η and σ_D.

reduction in risk-bearing capacity, gradual price declines occur over subsequent months as new capital finds its way into the sector. Chen, Noronha, and Singhal (2004), Greenwood (2005), Harris and Gurel (1986), and Kaul, Mehrotra, and Morck (2000), among others, document much shorter-lived price reactions to supply shocks in equity markets caused by changes in the definition of stock indices, at which time the position changes of committed index investors must be absorbed by other investors. Andrade, Chang, and Seasholes (2005) provide evidence of similar price reactions, again jumps followed by gradual price recoveries, following outside order imbalances on the Taiwanese stock market. While the stock markets treated in these studies are not over-the-counter, our model can be used as an abstraction of the process by which providers of liquid capital are located and drawn into the market in order to exploit temporary shortages in the levels of capital held by specialists. (Grossman and Miller (1988) and Weill (2003) focus only on

18

the supply of immediacy by specialists.)

While highly stylized, our model of periods of abnormal expected returns and price momemtum following supply shocks also provides additional microeconomic foundations for prior asset-pricing research on "limits to arbitrage," or "good deals," such as Shleifer and Vishny (1997) and Cochrane and Saa-Requejo (2001).

We adjust the baseline model of Section 1 (or, as explained, the linearized version of the risk-premium model of Section 2) by introducing occasional, randomly timed, aggregate liquidity shocks. At each such shock, a fraction of the agents, randomly chosen, suffer a sudden "reduction in liquidity," in the sense that their intrinsic types simultaneously jump to the low state. The shocks are timed according to a Poisson arrival process, independent of all other random variables, with mean arrival rate ζ.

Again appealing to the Law of Large Numbers, at each aggregate liquidity shock, the distribution of agents' intrinsic types becomes $\mu = \bar{\mu}$, where the post-shock distribution $\bar{\mu}$ is in $[0,1]^4$, satisfies (1)–(2), and has an abnormally elevated quantity of illiquid agents, both owners and non-owners. Specifically, $\bar{\mu}_{lo} > \mu_{lo}(t)$ and $\bar{\mu}_{ln} > \mu_{ln}(t)$. Each high-type owner remains a high type with probability $1 - \pi_{ho}(t) = \bar{\mu}_{ho}/\mu_{ho}(t)$, and becomes a low type with probability π_{ho}. Similarly, a high-type non-owner remains high type with probability $1 - \pi_{hn}(t) = \bar{\mu}_{hn}/\mu_{hn}(t)$ and becomes low type with probability π_{hn}. Conditional on $\pi(t)$, the changes in types are pairwise independent across the space of agents. This aggregate "liquidity shock" does not directly affect low-type agents. Of course, it affects them indirectly because of the change to the demographics of the market in which they live. By virtue of this specification, the post-shock distribution of agents does not depend on any residual "aftereffects" of prior shocks, a simplification without which the model would be relatively intractable.

In order to solve the equilibrium with an aggregate liquidity shock, it is helpful to use the "trick" of measuring time in terms of the passage of time t since the last shock, rather than absolute calendar time. Knowledge of the time at which this shock occurred enables an immediate translation of the solution into calendar time.

We first solve the equilibrium fractions $\mu(t) \in \mathbb{R}^4$ of agents of the four different types. At the time of an aggregate liquidity shock, this type distribution is equal to the post-shock distribution $\mu(0) = \bar{\mu}$ (where, to repeat, "0" means zero time units after the shock). After an aggregate liquidity shock, the cross-sectional distribution of agent types evolves according to the ODE

(3), converging (conditional on no additional shocks) to a steady state as the time since last shock increases.

Given this time-varying equilibrium solution of the investor type distribution, we turn to the agents' value functions. The value $V_\sigma(t)$ depends on the agent's type σ and the time t since the last aggregate liquidity shock. The values evolve according to

$$
\begin{aligned}
\dot{V}_{lo}(t) &= rV_{lo}(t) - \lambda_u(V_{ho}(t) - V_{lo}(t)) - 2\lambda\mu_{hn}(P(t) + V_{ln}(t) - V_{lo}(t)) \\
&\quad - \zeta(V_{lo}(0) - V_{lo}(t)) - (1 - \delta) \\
\dot{V}_{ln}(t) &= rV_{ln}(t) - \lambda_u(V_{hn}(t) - V_{ln}(t)) - \zeta(V_{ln}(0) - V_{ln}(t)) \\
\dot{V}_{ho}(t) &= rV_{ho}(t) - \lambda_d(V_{lo}(t) - V_{ho}(t)) \\
&\quad - \zeta((1 - \pi_{ho}(t))V_{ho}(0) + \pi_{ho}(t)V_{lo}(0) - V_{ho}(t)) - 1 \qquad (22) \\
\dot{V}_{hn}(t) &= rV_{hn} - \lambda_d(V_{ln} - V_{hn}) - 2\lambda\mu_{ho}(V_{ho} - V_{hn} - P) \\
&\quad - \zeta((1 - \pi_{hn}(t))V_{hn}(0) + \pi_{hn}(t)V_{ln}(0) - V_{hn}), \\
P(t) &= (V_{lo}(t) - V_{ln}(t))(1 - q) + (V_{ho}(t) - V_{hn}(t))q,
\end{aligned}
$$

where the terms involving ζ capture the risk of an aggregate liquidity shock. This differential equation is linear in the vector $V(t)$, depends on the deterministic evolution of $\mu(t)$, and has the somewhat unusual feature that it depends on the initial value function $V(0)$. To solve this system, it is useful to express it in the vector form:

$$
\dot{V}(t) = A_1(\mu(t))V(t) - A_2 - A_3(\mu(t))V(0), \qquad (23)
$$

where $A_1(\mu(t)) \in \mathbb{R}^{4\times4}$, $A_2 \in \mathbb{R}^{4\times1}$, and $A_3 \in \mathbb{R}^{4\times4}$ are the coefficient matrices. Treating $V(0)$ as a fixed parameter, the unique solution to the linear ODE that satisfies the appropriate transversality condition is

$$
V(t) = \int_t^\infty e^{-\int_t^s A_1(\mu(u))\,du} \left(A_2 + A_3(\mu(s))V(0)\right)\,ds. \qquad (24)
$$

At $t = 0$, this gives

$$
V(0) = \int_0^\infty e^{-\int_0^s A_1(\mu(u))\,du} \left(A_2 + A_3(\mu(s))V(0)\right)\,ds, \qquad (25)
$$

and we can then derive the initial value function $V(0)$ as the fixed point:

$$
V(0) = \left(I_4 - \int_0^\infty e^{-\int_0^s A_1(\mu(u))\,du} A_3(\mu(s))\,ds\right)^{-1} \left(\int_0^\infty e^{-\int_0^s A_1(\mu(u))\,du} A_2\,ds\right),
$$

$$\tag{26}$$

where $I_4 \in \mathbb{R}^{4 \times 4}$ is the identity matrix. Equations (24) and (26) together represent the solution. One notes that we take the bargaining power q as exogenous for simplicity, rather than incorporating the effects of delay during negotiation that stem for interim changes in the value functions.

Numerical Examples. We will illustrate some of the most noteworthy effects of a liquidity shock using a numerical example, and then state some general properties. We suppose that the search intensity is $\lambda = 125$, that types change idiosyncratically with intensities $\lambda_u = 2$ and $\lambda_d = 0.2$, that the fraction of owners is $s = 0.75$, that the riskless return is $r = 10\%$, that buyers and sellers have equal bargaining powers (that is, $q = 0.5$), that the illiquidity loss rate is $\delta = 2.5$, that the intensity of an aggregate liquidity shock is $\zeta = 0.1$, and that the post-shock distribution of types is determined by $\bar{\mu}_{lo} = 0.377$ and $\bar{\mu}_{ln} = 0.169$. These parameters are consistent with a shock from steady state[8] at which high types become low types with probability 0.5.

In order to motivate the results, one could imagine that an aggregate liquidity shock is associated with an event at which a large fraction of investors incur a significant loss of risk-bearing capacity, and thus have a higher shadow price for bearing risk. For example, at the default of Russia in 1998, those asset managers specializing in emerging market debt who had had positions in Russian issues would have had a substantially reduced appetite for holding Argentinian sovereign debt issues (the asset of concern), despite the lack of any material direct connection between the Russian and Argentinian economies, because of the asset managers' direct losses of capital due to the Russian default, and perhaps indirectly through demands for liquidation by relatively unsophisticated clients. Or, for example, when Hurricane Andrews struck, reinsurers with exposure to that event would have had a sudden reduction in capital available to cover monoline insurers facing, say, earthquake risk. While one could use the model of aversion to correlated endowment risk of Section 2 to compute the illiquidity loss rate δ associated with an aggregate shock, we would prefer to discuss the implications of the aggregate shock in more general terms, abstracting from the determination of the illiquidity loss rate δ.

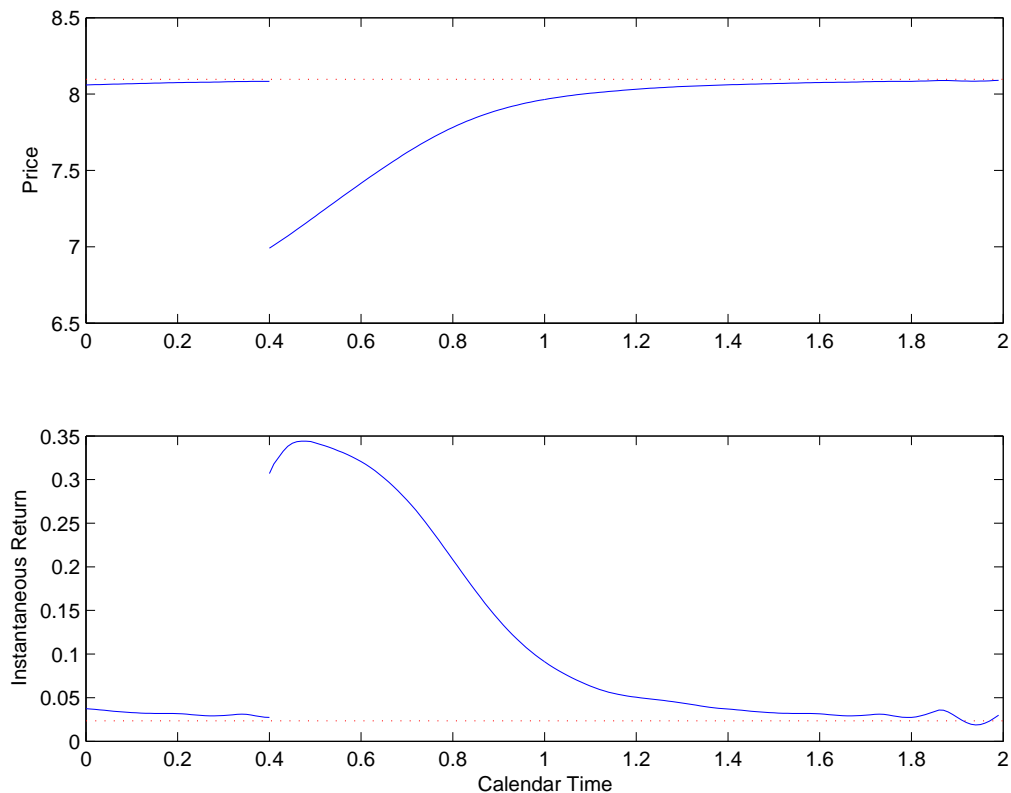[8]The steady-state masses, absent new shocks, are $\mu_{lo} = 0.004$ and $\mu_{ln} = 0.087$.

Figure 4: The top panel shows the price as a function of time when an aggregate liquidity shock at time 0.4. The bottom panel shows the corresponding annualized realized returns.

So, thinking of the aggregate shock as a simultaneous loss in capital to many investors that induces a reduced appetite by them for owning the asset in question, we may view an investor affected by the shock as having the intensity λ_u for a "recapitalizing" event, after which that investor no longer has an elevated "distress cost" δ for owning the asset.

The price and return dynamics associated with these parameters are shown in Figure 4. The top panel of the figure shows prices and the bottom panel shows realized instantaneous returns, both as functions of calendar time for a specific state of nature. At time $t = 0.4$, the economy experiences an aggregate liquidity shock, causing the asset price to drop suddenly by about 15%. Notably, it takes more than a year for the asset price to recover

to a roughly normal level. A buyer who was able to locate a seller immediately after the shock and realized an annualized return of roughly 30% for several months. While one is led to think in terms of the value to this "vulture" buyer of having retained "excess" liquidity so as to profit at times of aggregate liquidity shocks, our model has only one type of buyer, and is therefore too simple to address this sort of vulture specialization.

The large illustrated price impact of a shock is due to the large number of sellers and the relatively low number of potential buyers that are available immediately after a shock. The roughly 50% reduction in potential buyers that occurred at the illustrated shock increased a seller's expected time to find a potential buyer from 6.2 days immediately before the shock to 12.4 days immediately after the shock. Further, once a seller finds a buyer, the seller's bargaining position is poor because of his reduced outside search options and the buyer's favorable outside options.

Naturally, high-type owners who become low-type owners experience the largest utility loss from a shock.[9] The utility loss for low-type owners is also large, since their prospects of selling worsen significantly. High-type owners who do not themselves become low-type during the shock are not affected much since they expect the market to recover to normal conditions before they need to make a sale (given that $\lambda_d = 0.2$ is small relative to the length of the recovery period). The agents who don't hold the asset when the shock hits are positively affected since they stand a good chance of being able to benefit from the sell-side pressure.

The prospect of future aggregate liquidity shocks affects prices. For this, we compare the price a "long time after" a shock has hit (absent new shocks, that is, $\lim_{t\to\infty} P(t)$) with the steady-state asset price, $P^{\zeta=0} = 9.25$ associated with an economy with no aggregate shocks ($\zeta = 0$), but otherwise the same parameters. The presence of aggregate liquidity shocks reduces the price, in this sense, by 12.5%.

There are two channels through which the price responds after an aggregate shock. One is the search technology for trading. The other is the recovery of individual investors from the shock itself, captured by the switching intensities λ_u and λ_d. We have already motivated λ_u as the mean rate at which a financially distressed investor can be recapitalized.

---

[9]The utilities of the owners drop from $V_{ho} = 9.29$ and $V_{lo} = 9.14$, respectively, to $V_{ho} = 9.24$ and $V_{lo} = 8.47$, while the values of the non-owners increase from $V_{hn} = 1.13$ and $V_{ln} = 1.10$ to $V_{hn} = 2.22$ and $V_{ln} = 1.51$.

In order to better disentangle the effects of trading illiquidity from the effects of the shock on the distribution over time of advantaged (high-type) and disadvantaged agents, we change the numerical example so as to have more high-type investors than assets at all times, even after a shock. With perfectly liquid trading ($\lambda = +\infty$), therefore, the price would be unaffected by the aggregate liquidity shock. Specifically, we adjust the parameters used to create Figure 4 by reducing the probability that an individual agent is adversely affected by an aggregate shock from 0.5 to 0.17, so that $\bar{\mu}_{lo} = 0.128$ and $\bar{\mu}_{ln} = 0.115$. Figure 5 illustrates the time signature of the price reaction to an aggregate shock of this relatively benign variety, for two different values of the search intensity λ. Although the perfect-market price would be unaffected, search frictions cause an immediate negative return, followed by a price recovery over time that is accelerated by increasing the search intensity.

The features of these numerical examples are relatively general. An adverse liquidity shock causes an instantaneous price drop, price momentum during a relatively time-consuming recovery period, a reduced long-run price recovery level (due to the risk of future shocks), and an increase in expected selling times. The "time signature" of the price response reflects both the expected time for an adversely affected agent to recover (for example, for a distressed investor to find new financing), as captured by the parameter λ_u, and the time required for the assets to move from adversely affected sellers to potential buyers, in light of search frictions captured by the parameter λ. The latter effect incorporates both the trading delay due to search and the implications of temporarily superior outside options for potential buyers during negotiation with distressed sellers.

After an aggregate liquidity shock, the dynamics of agents' value functions and the price depend on the demography-induced time patterns of search times. In particular, a shock reduces the quantity of buyers and increases the quantity of sellers, which motivates part *a)* of the following proposition. Should the masses of these agents evolve monotonically following a shock, then so do the value functions and the price. It is possible, however, for the quantity of buyers to continue to decrease for some period after a shock before rebounding toward the steady-state value, due to the potentially large proportional increase in the quantity of sellers. The following proposition accounts for this and related effects.
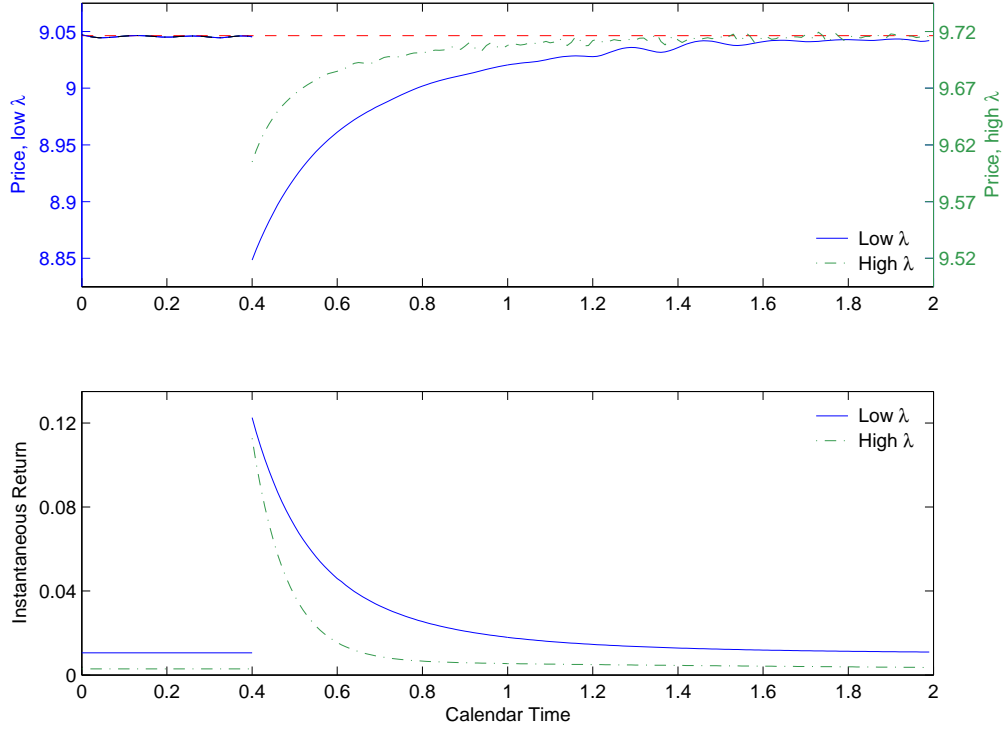
Figure 5: The top panel shows the price as a function of time when an aggregate liquidity shock happens at time 0.4. This is plotted both with a low search intensity $\lambda = 125$ (solid line, left axis), and a high search intensity $\lambda = 625$ (dashed lined, right axis). The bottom panel shows the annualized realized returns, again both for a low λ (solid line) and a high λ (dashed line).

Proposition 4 *There exist a time $\bar{T}$ and a strictly positive mean arrival rate $\bar{\zeta}$ such that:*

    a.) For any mean arrival rate ζ of aggregate shocks less than $\bar{\zeta}$ and time T larger than $\bar{T}$, if an aggregate shock arrives at T: (i) the jumps at T in the value functions of the owners (V_{ho} and V_{lo}) are downward, (ii) the jumps at T in the value functions of the non-owners (V_{hn} and V_{ln}) are upward, and (iii) the jump at T in the price is downward.

    b.) For any mean arrival rate ζ of aggregate shocks less than $\bar{\zeta}$ and any time t larger than $\bar{T}$: (i) the value functions of the owners (V_{ho} and V_{lo}) are increasing at t, (ii) the value functions of the non-owners (V_{hn}

25

and V_{ln}) are decreasing at t, and (iii) the price is increasing at t.

Moreover, if $\bar{\mu}$ is such that $\dot{\mu}_{hn}(0) \geq 0$ and $\dot{\mu}_{lo}(0) \leq 0$, then[10] one can take $\bar{T} = 0$.

4   Market Implications

We turn to a discussion of asset pricing in markets, particularly OTC markets, that are characterized by trade delayed by search for suitable counterparties. We discuss how our results are related to a variety of interesting empirical phenomena, and survey more generally the market implications of the recently growing literature on search in financial markets.

Exemplifying the imperfect ability to match buyers and sellers in OTC markets, traders in the market for European corporate loans ironically describe[11] trade in that market as "by appointment." Most major corporate bond markets are over-the-counter, and search problems are prevalent. Chen, Lesmond, and Wei (2005) relate this illiquidity of corporate bonds to their yield spread. Indeed, consistent with our Propositions 1 and 2, more illiquid bonds have higher yield spreads, that is, lower prices.

An extreme example of price impacts in relatively less liquid OTC markets is the study of Chinese equity prices by Chen and Xiong (2001). Certain Chinese companies have two classes of shares, one exchange traded, the other consisting of "restricted institutional shares" (RIS), which can be traded only privately. The two classes of shares are identical in every other respect, including their cash flows. Chen and Xiong (2001) find that RIS shares trade at an average discount of about 80% to the corresponding exchange-traded shares. Similarly, in a study involving U.S. equities, Silber (1991) compares the prices of "restricted stock" — which, for two years, can be traded only in private among a restricted class of sophisticated investors — with the prices of unrestricted shares of the same companies. Silber (1991) finds that restricted stocks trade at an average discount of 30%, and that the discount for restricted stock is increasing in the relative size of the issue. These price discounts can be captured in our search framework, but would be difficult to

---

[10]The equivalent conditions are that the mass dynamics due to trading are dominated by those due to the change in intrinsic types, namely $\lambda_d \bar{\mu}_{ho} - \lambda_u \bar{\mu}_{lo} \leq 2\lambda \bar{\mu}_{hn} \bar{\mu}_{lo}$, respectively $2\lambda \bar{\mu}_{hn} \bar{\mu}_{lo} \leq \lambda_u \bar{\mu}_{ln} - \lambda_d \bar{\mu}_{hn}$.

[11]See, for example, *The Financial Times*, November 19, 2003.

explain using standard models based on asymmetric information, given that the two classes of shares are claims to the same dividend streams, and given that the publicly-traded share prices are easily observable.

A large literature, surveyed by Amihud, Mendelson, and Pedersen (2006), addresses liquidity premia in equity markets, focusing mainly on non-search sources of illiquidity. Lagos (2005) examines the contribution to the equity premium puzzle of illiquidity in a search-based asset pricing model.

Even in the most liquid OTC markets the relatively small price effects arising from search frictions receive significant attention by economists. For example, the market for U.S. Treasury securities, an over-the-counter market considered to be a benchmark for high liquidity, has widely noted illiquidity effects that differentiate the yields of on-the-run (latest-issue) securities from those of off-the-run securities. Positions in on-the-run securities are normally available in large amounts from relatively easily found traders such as hedge funds and government-bond dealers. Because on-the-run issues can be more quickly located by short-term investors such as hedgers and speculators, they command a price premium, even over a package of off-the-run securities of identical cash flows. Ironically, episodically large on-the-run premia could actually be partly due to "scarcity premia," in the sense of Section 1. That is, because their superior liquidity causes some on-the-run issues to be such a dominant vehicle for trade, the extremely high velocity of circulation demanded by market participants can at some times be large relative to that available in light of the limits to the OTC search technology, and small but notable price premia can apply. The importance ascribed to these relatively small premia is explained by the exceptionally high volume of trade in this market, and also by the importance of disentangling the illiquidity impact on measured Treasury interest rates for informational purposes elsewhere in the economy. Longstaff (2002) measures relatively larger illiquidity effects on government security prices during "flights to liquidity," which he characterizes as periods during which a large demand for quick access to a safe haven causes Treasury prices to temporarily achieve markedly higher prices than equally safe government securities that are not as easily found.

Part of the spread between on-the-run and off-the-run treasuries is due to a premium in the effective lending fees for on-the-run issues that is larger when on-the-run issues are higher to find. In the OTC equities lending market, traders use terminology such as "to get a locate" of lendable shares. A search-based theory of securities lending is developed in Duffie, Gârleanu, and Pedersen (2002) and extended to multiple assets in Vayanos and Weill (2005).

27

Empirical evidence of the impact on treasury prices and securities-lending premia ("repo specials") can be found in Duffie (1996), Jordan and Jordan (1997), and Krishnamurthy (2002), who estimates that much of the on-the-run price premia in 30-year issues has been due, on average, to repo specials. Lending "specials" in equity markets are measured by Geczy, Musto, and Reed (2002), D'Avolio (2002), and Jones and Lamont (2002). Difficulties in locating lenders of shares sometimes cause dramatic price imperfections, as was the case with the spinoff of Palm, Incorporated, one of a number of such cases documented by Mitchell, Pulvino, and Stafford (2002). Fleming and Garbade (2003) document a new U.S. Government program to improve liquidity in treasury markets by allowing alternative types of treasury securities to be deliverable in settlement of a given repurchase agreement, mitigating the costs of search for a particular issue.

Consistent with the results of Section 2, that search frictions exacerbate risk premia stemming from hedging motives, Graveline and McBrady (2005) find that Treasury repo specials are empirically linked to hedging, particularly by financial firms exposed to inventories of mortgage-backed securities and corporate bonds. In particular, repo specials are higher when the inventories to be hedged are larger, and when interest-rate volatility is higher.

Section 3 shows how our model can be used to characterize the implications of a widespread shock to the abilities or incentives of traders to take asset positions. An increase in the number of would-be sellers and a reduction in the number of potential buyers result in a price drop, in part because of the higher fraction of assets held by distressed traders, and also because of the worsened bargaining position of sellers. Over time, the price recovers as distressed sellers recover from the adverse effects of the shock itself, and as trading, limited by search frictions, reallocates the asset from distressed sellers to potential buyers.

One finds this sort of time signature of price reactions to supply or demand shocks in several markets. For instance, in the corporate bond markets, one observes large price drops and delayed recovery in connection with major downgrades or default (Hradsky and Long (1989)), when certain classes of investors have an incentive (or a contractual requirement) to sell their holdings. Anecdotally, similar reactions in the prices of emerging-market sovereign debt frequently occur, for example, during a major debt crisis (though it is hard to measure "fundamentals" in this case). Another prominent example is provided by the reinsurance markets, in which such price patterns occur in response to sudden large aggregate claims against providers of insurance at

times of major natural disasters, as documented by Froot and O'Connell (1999).

As an example of the impact of aggregate liquidity shocks described by Proposition 4, when stocks are "dumped" by index investors during index recomposition events, the associated price reactions, as described by Shleifer (1986), Harris and Gurel (1986), Kaul, Mehrotra, and Morck (2000), Chen, Noronha, and Singhal (2004), Greenwood (2005), among others, have time signatures of the sort predicted by our model.

The dynamics of price reactions to outside order imbalances, documented for example in Andrade, Chang, and Seasholes (2005) and Coval and Stafford (2005) are also consistent with the thrust of our model in Section 3. While Andrade, Chang, and Seasholes (2005) analyze a specialist market rather than an OTC market, they indicate that the most apparent suppliers of liquidity to this market at times of order imbalances are outside investors, who seem to take time, even days, to "find" the trading opportunity in sufficient numbers to drive the prices back to normal. Coval and Stafford (2005) document analogous stock-price reactions associated with imbalanced inflows and redemptions of capital to mutual funds. Weill (2003) characterizes the optimal behavior of marketmakers in absorbing supply shocks in order to mitigate search frictions by "leaning against" the outside order flow.

As we show in Section 2, higher risk (in the form of higher dividend volatility or higher correlation between the dividends and an agent's endowed income) magnifies the impact on prices of search frictions, because higher risk increases the utility losses of distressed owners until they can find a suitable buyer. One can generalize the model (as we did in a previous version of the paper) to incorporate risk limits on the size of traders' positions. Agents compensate for larger risks with reduced position limits. As a result, a larger fraction of agents must hold the risky asset, and liquidity is further reduced because finding a buyer is more difficult. Hence, shocks to volatility can lead to increased liquidity frictions and price drops, especially if risk-management practices imply a simultaneous tightening of position limits.

Search frictions also help explain how the relative size of an asset in the economy may affect its price (or price-dividend ratio). Proposition 1 of this paper shows that if a higher fraction (s) of the agents must hold the asset, then the price must fall. When comparing stocks cross-sectionally, moreover, a role for search frictions arises from the additional implications of selection of participants into a market. Investors prefer, all else equal, to participate in markets for larger stocks. If, for instance, the number of

investors participating in the market for a firm's shares is proportional to the size of the firm, correspondingly shorter mean delays for locating suitable counterparties for trades in the shares of larger firms would lead to higher price-dividend ratios. Such cross-sectional asset-pricing results are studied more directly by Weill (2002) and Vayanos and Wang (2002), who extended our baseline model to treat multiple assets and showed, among other results, that securities with a larger free float (shares available for trade) are more liquid and have lower expected returns, and that concentrations of trade in a favored security, may explain some of the price difference between on-the-run and off-the-run Treasury securities.

Duffie, Gârleanu, and Pedersen (2005) study the implications of search frictions for marketmakers.[12] Here, outside investors remain able to find other outside investors with some search intensity λ, but can also find marketmakers with some intensity ρ. This framework captures the feature that investors bargain sequentially with marketmakers. The price negotiation between a marketmaker and an investor reflects the investor's outside options, including in particular the investor's ability to meet and trade with other investors or marketmakers. It is shown that the marketmaker's bid-ask spread is lower if the investor can find other investors on his own more easily. Further, the spread is lower if an investor can approach other marketmakers more easily. In other words, more "sophisticated" investors are quoted *tighter* spreads by marketmakers. Examples can be found in the typical hub-and-spoke structure of contact among marketmakers and their customers in OTC derivative markets. This distinguishes our search-based theory from traditional information-based theories that predict that more sophisticated (in this setting, more informed) investors are quoted *wider* spreads by marketmakers (Glosten and Milgrom (1985)).

In OTC markets for interest-rate derivatives, a "sales trader" and an outside customer negotiate a price, implicitly including a dealer profit margin that is based in part on the customer's (perceived) outside option. The risk that customers have superior information about future interest rates is normally regarded as small. The customer's outside option depends on how easily he can find a counterparty himself and how easily he can access other dealers. As explained by Commissioner of Internal Revenue (2001) (page 13) in recent litigation regarding the portion of dealer margins on interest-rate

---

[12]Other search-based models of intermediation include Rubinstein and Wolinsky (1987), Bhattacharya and Hagerty (1987), Moresi (1991), Gehrig (1993), and Yavaş (1996).

swaps that can be ascribed to dealer profit, dealers typically negotiate prices with outside customers that reflect the customer's relative lack of access to other market participants. In order to trade OTC derivatives with a bank, for example, a customer must have, among other arrangements, an account and a credit clearance. Smaller customers often have an account with only one, or perhaps a few, banks, and therefore have fewer search options. Hence, a testable implication of a version of this model with investors of heterogeneous search intensities is that investors with fewer search options (typically, small unsophistcated investors) receive less competitive prices. We note that these "small" investors are less likely to be informed, so that models based on informational asymmetries alone would reach the opposite prediction.

Newman and Rierson (2003) use our approach in a search-based model of corporate bond pricing, in which large issues of credit-risky bonds temporarily raise credit spreads throughout the issuer's sector, because providers of liquidity such as underwriters and hedge funds bear extra risk as they search for long-term investors. They provide empirical evidence of temporary bulges in credit spreads across the European Telecom debt market during 1999-2002 in response to large issues by individual firms in this sector. Studying a different set of markets, Mikkelson and Partch (1985) find empirical support for "the notion that underwriting spreads are in part compensation for the selling effort." In particular, they find that underwriting spreads are positively related to the size of the offering.

Appendices

A   Explicit Bargaining Game

The setting considered here is that of Section 1, with two exceptions. First, agents can interact only at discrete moments in time, Δ_t apart. Later, we return to continuous time by letting Δ_t go to zero. Second, the bargaining game is modeled explicitly.

We follow Rubinstein and Wolinsky (1985) and others in modeling an alternating-offers bargaining game, making the adjustments required by the specifics of our setup. When two agents are matched, one of them is chosen randomly — the seller with probability $\hat{q}$, the buyer with probability $1 - \hat{q}$ — to suggest a trading price. The other either rejects or accepts the offer,

immediately. If the offer is rejected, the owner receives the dividend from the asset during the current period. At the next period, Δ_t later, one of the two agents is chosen at random, independently, to make a new offer. The bargaining may, however, break down before a counteroffer is made. A breakdown may occur because either of the agents changes valuation type, whence there are no longer gains from trade. A breakdown may also occur if one of the agents meets yet another agent, and leaves his current trading partner. The latter reason for breakdown is only relevant if agents are allowed to search while engaged in negotiation.

We consider first the case in which agents can search while bargaining. We assume that, given contact with an alternative partner, they leave the present partner in order to negotiate with the newly found one. The offerer suggests the price that leaves the other agent indifferent between accepting and rejecting it. In the unique subgame perfect equilibrium, the offer is accepted immediately (Rubinstein (1982)). The value from rejecting is associated with the equilibrium strategies being played from then onwards. Letting P_σ be the price suggested by the agent of type σ with $\sigma \in \{lo, hn\}$, letting $\bar{P} = \hat{q}P_{lo} + (1 - \hat{q})P_{hn}$, and making use of the motion laws of V_{lo} and V_{hn}, we have

$$
\begin{aligned}
P_{hn} - \Delta V_l &= e^{-(r+\lambda_d+\lambda_u+2\lambda\mu_{lo}+2\lambda\mu_{hn})\Delta_t}(\bar{P} - \Delta V_l) + O(\Delta_t^2) \\
-P_{lo} + \Delta V_h &= e^{-(r+\lambda_d+\lambda_u+2\lambda\mu_{lo}+2\lambda\mu_{hn})\Delta_t}(-\bar{P} + \Delta V_h) + O(\Delta_t^2).
\end{aligned}
$$

These prices, P_{hn} and P_{lo}, have the same limit $P = \lim_{\Delta_t \to 0} P_{hn} = \lim_{\Delta_t \to 0} P_{lo}$. The two equations above readily imply that the limit price and limit value functions satisfy

$$
P = \Delta V_l (1 - q) + \Delta V_h q, \tag{A.1}
$$

with

$$
q = \hat{q}. \tag{A.2}
$$

This result is interesting because it shows that the seller's bargaining power, q, does not depend on the parameters — only on the likelihood that the seller is chosen to make an offer. In particular, an agent's intensity of meeting other trading partners does not influence q. This is because one's own ability to meet an alternative trading partner: *(i)* makes oneself more impatient, and *(ii)* also increases the partner's risk of breakdown, and these two effects cancel out.

This analysis shows that the bargaining outcome used in our model can be justified by an explicit bargaining procedure. We note, however, that other bargaining procedures lead to other outcomes. For instance, if agents cannot search for alternative trading partners during negotiations, then the same price formula (A.1) applies with

$$q = \frac{\hat{q}(r + \lambda_u + \lambda_d + 2\lambda\mu_{lo})}{\hat{q}(r + \lambda_u + \lambda_d + 2\lambda\mu_{lo}) + (1 - \hat{q})(r + \lambda_u + \lambda_d + 2\lambda\mu_{hn})}. \qquad (A.3)$$

This bargaining outcome would lead to a similar solution for prices, but the comparative-static results would change, since the bargaining power q would depend on the other parameters.

B Walrasian Equilibrium with Risk Aversion

This section derives the competitive equilibrium with risk averse agents (as in Section 2) who can immediately trade *any* number of risky securities. We note that this is different from a competitive market with fixed exogenous position sizes, that is, it is different from the limit considered in Proposition 3.

Suppose that the Walrasian price is constant at P, that is, agents can trade instantly at this price. An agent's total wealth — cash plus the value of his position in risky assets — is denoted by $\bar{W}$. If an agent chooses to hold $\theta(t)$ shares at any time t, then the wealth-dynamics equation is

$$d\bar{W}_t = (r\bar{W}_t - r\theta_t P - c_t)\, dt + \theta_t\, dD_t + d\eta_t.$$

The HJB equation for an agent of intrinsic type $\sigma \in \{h, l\}$ is

$$\begin{aligned}
0 &= \sup_{\bar{c}, \theta} \{J_w(w, \sigma)(rw - \bar{c} + \theta(m_D - rP) + m_\eta) \\
&\quad + \frac{1}{2}J_{ww}(w, \sigma)(\theta^2 \sigma_D^2 + \sigma_\eta^2 + 2\rho_\sigma \theta \sigma_D \sigma_\eta) \\
&\quad + \lambda(\sigma, \sigma')[J(w, \sigma) - J(w, \sigma')] - e^{-\gamma\bar{c}} - \beta J(w, \sigma)\},
\end{aligned}$$

where $\lambda(\sigma, \sigma')$ is the intensity of change of intrinsic type from σ to σ'. Conjecturing the value function $J(w, \sigma) = -e^{-r\gamma(w + a_\sigma + \bar{a})}$, optimization over θ yields

$$\theta_\sigma = \frac{m_D - rP - r\gamma\rho_\sigma\sigma_D\sigma_\eta}{r\gamma\sigma_D^2}. \qquad (B.1)$$

Market clearing requires

$$\mu_h \theta_h + \mu_l \theta_l = \Theta,$$

with $\mu_h = 1 - \mu_l = \lambda_u/(\lambda_u + \lambda_d)$, which gives the price

$$P = \frac{m_D}{r} - \gamma \left(\Theta \sigma_D^2 + \frac{\sigma_D \sigma_\eta \left[\rho_l \lambda_d + \rho_h \lambda_u \right]}{\lambda_u + \lambda_d} \right). \tag{B.2}$$

Inserting this price into (B.1) gives the quantity choices

$$\theta_h = \Theta + \frac{\sigma_\eta \lambda_d \left[\rho_l - \rho_h \right]}{\sigma_D (\lambda_u + \lambda_d)} \tag{B.3}$$

$$\theta_l = \Theta - \frac{\sigma_\eta \lambda_u \left[\rho_l - \rho_h \right]}{\sigma_D (\lambda_u + \lambda_d)}. \tag{B.4}$$

C Proofs

Proof of Proposition 1: The dependence on δ and q is seen immediately, given that no other variable entering Equation (7) depends on either δ or q.

Viewing P and μ_σ as functions of the parameters λ_d and s, a simple differentiation exercise shows that the derivative of the price P with respect to λ_d is a positive multiple of

$$(rq + \lambda_u + 2\lambda \mu_{hn} q) \left(1 + 2\lambda \frac{\partial \mu_{lo}}{\partial \lambda_d} (1 - q) \right)$$
$$- \ (r(1 - q) + \lambda_d + 2\lambda \mu_{lo}(1 - q)) \left(2\lambda \frac{\partial \mu_{hn}}{\partial \lambda_d} q \right),$$

which is positive if $\frac{\partial \mu_{lo}}{\partial \lambda_d}$ is positive and $\frac{\partial \mu_{hn}}{\partial \lambda_d}$ is negative.

These two facts are seen as follows. From Equations (1)-(3) and the fact that $\mu_{lo} + \mu_{ln} = \lambda_d (\lambda_d + \lambda_u)^{-1} = 1 - y$, where

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

it follows that μ_{lo} solves the equation

$$2\lambda \mu_{lo}^2 + (2\lambda(y - s) + \lambda_u + \lambda_d)\mu_{lo} - \lambda_d s = 0. \tag{C.1}$$

This quadratic equation has a negative root and a root in the interval $(0, 1)$, and this latter root is μ_{lo}.

Differentiating (C.1) with respect to λ_d, one finds that

$$\frac{\partial \mu_{lo}}{\partial \lambda_d} = \frac{s - \mu_{lo} - 2\lambda \frac{\partial y}{\partial \lambda_d} \mu_{lo}}{2\lambda \mu_{lo} + 2\lambda(y - s) + \lambda_u + \lambda_d} > 0,$$

since $\frac{\partial y}{\partial \lambda_d} < 0$. Similar calculations show that

$$\frac{\partial \mu_{hn}}{\partial \lambda_d} = \frac{-\lambda_d + 2\lambda \frac{\partial y}{\partial \lambda_d} \mu_{hn}}{2\lambda \mu_{lo} + \lambda_u + \lambda_d} < 0,$$

which ends the proof of the claim that the price decreases with λ_d. Like arguments can be used to show that $\frac{\partial \mu_{lo}}{\partial \lambda_u} < 0$ and that $\frac{\partial \mu_{hn}}{\partial \lambda_u} > 0$, which implies that P increases with λ_u.

Finally,

$$\frac{\partial \mu_{lo}}{\partial s} = \frac{\lambda_d + 2\lambda \mu_{lo}}{2\lambda \mu_{lo} + 2\lambda(y - s) + \lambda_u + \lambda_d} > 0$$

and

$$\frac{\partial \mu_{hn}}{\partial s} = \frac{-\lambda_u - 2\lambda \mu_{hn}}{2\lambda \mu_{lo} + \lambda_u + \lambda_d} < 0,$$

showing that the price decreases with the supply s.

In order to prove that the price increases with λ for λ large enough, it is sufficient to show that the derivative of the price with respect to λ changes sign at most a finite number of times, and that the price tends to its upper bound, $1/r$, as λ tends to infinity. The first statement is obvious, while the second one follows from Equation (7), given that, under the assumption $s < \lambda_u/(\lambda_u + \lambda_d)$, $\lambda \mu_{lo}$ stays bounded and $\lambda \mu_{hn}$ goes to infinity with λ.

$\square$

Proof of Proposition 2: We impose on investors' choices of consumption and trading strategies the transversality condition that, for any initial agent type σ_0, $e^{-\beta T} E_0[e^{-r\gamma W_T}] \to 0$ as T goes to infinity. Intuitively, the condition means that agents cannot consume large amounts forever by increasing their debt without restriction. We must show that our candidate optimal consumption and trading strategy satisfies that condition.

We conjecture that, for our candidate optimal strategy, $E_0[J(W_T, \sigma_T)] = e^{(\beta-r)T}J(W_0, \sigma_0)$. Clearly, this implies that the transversality condition is satisfied, since

$$
\begin{aligned}
e^{-\beta T}E_0[e^{-r\gamma W_T}] &= -e^{-\beta T}E_0\left[J(W_T, \sigma_T)e^{r\gamma(\bar{a}+a_{\sigma_T})}\right] \\
&\leq -\sup_\sigma e^{r\gamma(\bar{a}+a_\sigma)}e^{-rT}J(W_0, \sigma_0) \\
&\to 0.
\end{aligned}
$$

This conjecture is based on the insights that *(i)* the marginal utility, $u'(c_0)$, of time-0 consumption must be equal to the marginal utility, $e^{(r-\beta)T}E_0[u'(c_T)]$, of time T consumption; and *(ii)* the marginal utility is proportional to the value function in our (CARA) framework. (See Wang (2002) for a similar result.)

To prove our conjecture, we consider, for our candidate optimal policy, the wealth dynamics

$$
\begin{aligned}
dW &= \left(\frac{\log r}{\gamma} - ra_\sigma - r\bar{a} + \theta_\sigma m_D + m_\eta\right)dt + \theta_\sigma \sigma_D\, dB + \sigma_\eta\, dB^i - P\, d\theta_\sigma \\
&= \left(-ra_\sigma + \theta_\sigma m_D + \frac{1}{2}r\gamma\sigma_\eta^2 + \frac{r-\beta}{r\gamma}\right)dt + \theta_\sigma \sigma_D\, dB + \sigma_\eta\, dB^i - P\, d\theta_\sigma \\
&= M(\sigma)\, dt + \sqrt{\Sigma(\sigma)}\, d\hat{B} - P\, d\theta_\sigma,
\end{aligned}
$$

where M, Σ and the standard Brownian motion $\hat{B}$ are defined by the last equation.

Define f by

$$
f(W_t, \sigma_t, t) = E_t[J(W_T, \sigma_T)] = -E_t[e^{-r\gamma(W_T + a_{\sigma_T} + \bar{a})}].
$$

Then, by Ito's Formula,

$$
\begin{aligned}
0 &= f_t + f_w M(\sigma) + \frac{1}{2}f_{ww}\Sigma(\sigma) \qquad\qquad\qquad\qquad\qquad (C.2) \\
&\quad + \sum_{\{\sigma':\sigma'\neq\sigma\}} \lambda(\sigma, \sigma')\left(f(w + z(\sigma, \sigma')P, \sigma', t) - f(w, \sigma, t)\right),
\end{aligned}
$$

where $\lambda(\sigma, \sigma')$ is the intensity of transition from σ to σ' and $z(\sigma, \sigma')$ is -1, 1, or 0, depending on whether the transition is, respectively, a buy, a sell, or an intrinsic-type change. The boundary condition is $f(w, \sigma, T) = -e^{-r\gamma(w + a_\sigma + \bar{a})}$.

The fact that $f(w, \sigma, t) = e^{(\beta-r)(T-t)}J(w, \sigma)$ now follows from the facts that *(i)* this function clearly satisfies the boundary condition, and *(ii)* it solves (C.2), which is confirmed directly using (16) for a_σ.

$\square$

Proof of Proposition 3: This result follows from Equations (16)–(19) as well as the fact that $\lambda\mu_{hn} \to \infty$ and $\lambda\mu_{lo}$ is bounded.

$\square$

Proof of Proposition 4: We start by considering two versions of the system (22), characterized by masses $\mu_{hn}^i(t)$ and $\mu_{lo}^i(t)$, $i \in \{1,2\}$, such that $\mu_{hn}^1(t) \geq \mu_{hn}^2(t)$, $\mu_{lo}^1(t) \leq \mu_{lo}^2(t)$, $\pi_{ho}^1 \geq \pi_{ho}^2$, and $\pi_{hn}^1 \geq \pi_{hn}^2$. We show that, if ζ is small enough, the following relationships hold at all times t:

$$
\begin{aligned}
V_{lo}^1 &\geq V_{lo}^2 \\
V_{ho}^1 &\geq V_{ho}^2 \\
V_{hn}^1 &\leq V_{hn}^2 \\
V_{ln}^1 &\leq V_{ln}^2 \\
P^1 &\geq P^2.
\end{aligned}
\tag{C.3}
$$

We then use this result to prove the Proposition.

Letting $\Delta V_o = V_{ho} - V_{lo}$, $\Delta V_n = V_{hn} - V_{ln}$, and $\phi = \Delta V_o - \Delta V_n$, the motion equations (22) for $\zeta = 0$, for any of the two mass configurations, are

$$
\begin{aligned}
\dot{V}_{lo} &= (r+\zeta)V_{lo} - \lambda_u\Delta V_o - 2\lambda\mu_{hn}q\phi - \zeta V_{lo}(0) - (1-\delta) & \text{(C.4)} \\
\dot{V}_{ln} &= (r+\zeta)V_{ln} - \lambda_u\Delta V_n - \zeta V_{ln}(0) & \text{(C.5)} \\
\dot{V}_{ho} &= (r+\zeta)V_{ho} + \lambda_d\Delta V_o - \zeta V_{ho}(0) + \zeta\pi_{ho}\Delta V_o(0) - 1 & \text{(C.6)} \\
\dot{V}_{hn} &= (r+\zeta)V_{hn} + \lambda_d\Delta V_n - 2\lambda\mu_{ho}(1-q) - \zeta V_{hn}(0) & \text{(C.7)} \\
&\quad + \zeta\pi_{ho}\Delta V_n(0)\phi \\
\Delta\dot{V}_o &= (r+\lambda_d+\lambda_u+\zeta)\Delta V_o + 2\lambda\mu_{hn}q\phi & \text{(C.8)} \\
&\quad - \zeta(1-\pi_{ho})\Delta V_o(0) - \delta \\
\Delta\dot{V}_n &= (r+\lambda_d+\lambda_u+\zeta)\Delta V_n - 2\lambda\mu_{lo}(1-q)\phi & \text{(C.9)} \\
&\quad - \zeta(1-\pi_{hn})\Delta V_n(0) \\
\dot{\phi} &= (r+\lambda_d+\lambda_u+2\lambda\mu_{hn}q+2\lambda\mu_{lo}(1-q)+\zeta)\phi - \delta & \text{(C.10)} \\
&\quad - \zeta(1-\pi_{ho})\Delta V_o(0) + \zeta(1-\pi_{hn})\Delta V_n(0)
\end{aligned}
$$

Letting $\psi = (\Delta V_o, \Delta V_n)^\top$, Equations (C.8)–(C.9) can be further written as

$$
\dot{\psi} = A\psi + B,
\tag{C.11}
$$

37

with

$$A = \begin{bmatrix} r + \lambda_d + \lambda_u + \zeta + 2\lambda\mu_{hn}q & -2\lambda\mu_{hn}q \\ -2\lambda\mu_{lo}(1-q) & r + \lambda_d + \lambda_u + \zeta + 2\lambda\mu_{lo}(1-q) \end{bmatrix}$$

and

$$B = -\begin{bmatrix} \zeta(1-\pi_{ho})\Delta V_o(0) + \delta \\ \zeta(1-\pi_{hn})\Delta V_n(0) \end{bmatrix}.$$

We are going to show that, given the way in which the entries of A^1 and B^1 compare with those of A^2, respectively B^2, $\psi^1 \leq \psi^2$ for all t. To that end, consider a continuum of systems defined by $A_\alpha(t) = \alpha A^1(t) + (1-\alpha)A^2(t)$ and $B_\alpha(t) = \alpha B^1(t) + (1-\alpha)B^2(t)$, and consider the derivative of (C.11) with respect to α,

$$\frac{\partial}{\partial t}\left(\frac{\partial \psi_\alpha}{\partial \alpha}\right) = \frac{\partial A_\alpha}{\partial \alpha}\psi + A_\alpha\frac{\partial \psi_\alpha}{\partial \alpha} + \frac{\partial B_\alpha}{\partial \alpha}. \tag{C.12}$$

The solution is

$$\frac{\partial \psi_\alpha}{\partial \alpha}(t) = -\int_t^\infty e^{-\int_t^s A_\alpha(u)\,du}\left(\frac{\partial A_\alpha}{\partial \alpha} + \frac{\partial B_\alpha}{\partial \alpha}\right)ds. \tag{C.13}$$

Note that

$$\frac{\partial A_\alpha}{\partial \alpha} + \frac{\partial B_\alpha}{\partial \alpha} = 2\lambda\phi_\alpha\begin{bmatrix} q\,(\mu_{hn}^1 - \mu_{hn}^2) \\ (1-q)\,(\mu_{lo}^2 - \mu_{lo}^1) \end{bmatrix} + \zeta\begin{bmatrix} (\pi_{ho}^1 - \pi_{ho}^2)\Delta V_o(0) \\ (\pi_{hn}^1 - \pi_{hn}^2)\Delta V_n(0) \end{bmatrix}$$

is positive, so it suffices to sign the elements of the matrix e^M, where the matrix M is of the form

$$M = \begin{bmatrix} c - a & a \\ b & c - b \end{bmatrix}$$

with $a > 0$ and $b > 0$. It is immediate that the signs of the elements of e^M are the same as for the matrix $e^K e^M$ for any scalar K — in particular, for K large enough to make all entries of $K + M$ positive. We conclude that $\partial\psi/\partial\alpha \leq 0$, so that $\psi^1 \leq \psi^2$.

It now follows from (C.5) that $V_{ln}^1 \leq V_{ln}^2$, which, together with $\Delta V_n^1 \leq \Delta V_n^2$, implies that $V_{hn}^1 \leq V_{hn}^2$. The matters are not as simple with the owner value functions, since the fact that a positive ζ makes the chances of being hit by a shock lower for parameter set 2 works against our desired conclusion. This is why we need ζ small enough.

In order to conclude that $V_{ho}^1 \geq V_{ho}^2$, we need that

$$\lambda_d \left(\Delta V_o^2 - \Delta V_o^1 \right) - \zeta \left(\pi_{ho}^1 - \pi_{ho}^2 \right) \Delta V_o(0) \geq 0$$

for all t. Since the inequality holds for $\zeta = 0$, it suffices to show that

$$\lim_{t \to \infty} \frac{\Delta V_o^2 - \Delta V_o^1}{\pi_{ho}^1 - \pi_{ho}^2} > 0$$

and make use of continuity. Equation (C.13) provides a desired lower bound on $\Delta V_o^2 - \Delta V_o^1$ of at least $\epsilon \left(\pi_{ho}^1 - \pi_{ho}^2 \right)$ for t large enough and $\epsilon > 0$ small enough.

Having obtained that $V_{ho}^1 \geq V_{ho}^2$, it immediately follows from the definition of ΔV_o that $V_{lo}^1 \geq V_{lo}^2$. Finally, it is clear that $P^1 \geq P^2$.

We use the comparison result just proved to prove the Proposition. We start with part $a.$). By letting $\mu^2(t) = \mu(t)$, i.e., the economy under consideration, and $\mu^1 = \mu(\infty)$, i.e., a steady-state economy without shocks, it follows that the values of owners and the price are strictly lower at any point following the shock — in particular, at 0 — than their eventual value. The opposite is true for the values of the non-owners. To complete the proof of this statement, we need to show that, for all t, $\mu_{hn}(t) < \mu_{hn}(\infty)$ and $\mu_{lo}(t) > \mu_{lo}(\infty)$.

To that end, consider the dynamics of μ_{lo}:

$$\dot{\mu}_{lo}(t) \;=\; -2\lambda \mu_{lo}(t)^2 - \left(2\lambda \left(\mu_h(t) - s \right) + \lambda_u + \lambda_d \right) \mu_{lo}(t) + \lambda_d s.$$

Since $\mu_h(t) < \mu_h(\infty)$, $\mu_{lo}(0) > \mu_{lo}(\infty)$, and $\mu_{lo}(t) > 0$ a simple comparison theorem (e.g., Birkhoff and Rota (1969), page 25) shows that $\mu_{lo}(t) > \mu_{lo}(\infty)$ for all t. Likewise, $\mu_{hn}(t) < \mu_{hn}(\infty)$.

For part $b.$), let $\mu^2(t) = \mu(t)$ and $\mu^1(t) = \mu(t + dt)$ for an arbitrary $dt > 0$ and monotonicity of the value functions and price with respect to time follows, provided that the masses are monotonic. The masses, however, need not be monotonic. In general, only one of μ_{lo} and μ_{hn} is monotonic, while the other may move away from its steady-state value for a while before its derivative changes sign. Both masses are therefore monotonic for all t if $\dot{\mu}_{hn}(0) \geq 0$ and $\dot{\mu}_{lo}(0) \leq 0$, which proves the last assertion of the Proposition. Otherwise, they are monotonic for $t \geq T_2$ for some $T_2 > 0$.

$\square$

References

Acharya, V. and L. H. Pedersen (2005). Asset Pricing with Liquidity Risk. *Journal of Financial Economics 77*, 375–410.

Amihud, Y. and H. Mendelson (1986). Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics 17*, 223–249.

Amihud, Y., H. Mendelson, and L. H. Pedersen (2006). Liquidity and Asset Prices. *Foundations and Trends in Finance, forthcoming*.

Andrade, G., C. Chang, and M. Seasholes (2005). Predictable Reversals, Cross-Stock Effects, and the Limits of Arbitrage. Working Paper, Haas School of Business, University of California.

Bhattacharya, S. and K. M. Hagerty (1987). Dealerships, Trading Externalities, and General Equilibrium. In E. Prescott and N. Wallace (Eds.), *Contractual Arrangements for Intertemporal Trade*, Chapter Minnesota Studies in Macroeconomics Series, vol. 1, pp. 81–104. Minneapolis: University of Minnesota Press.

Birkhoff, G. and G.-C. Rota (1969). *Ordinary Differential Equations*. John Wiley & Sons.

Chen, H., G. Noronha, and V. Singhal (2004). The Price Response to S&P 500 Index Additions and Deletions: Evidence of Asymmetry and a New Explanation. *Journal of Finance 59*, 1901–1929.

Chen, L., D. Lesmond, and J. Wei (2005). Corporate Yield Spreads and Bond Liquidity. *Journal of Finance, forthcoming*.

Chen, Z. and P. Xiong (2001). Discounts On Illiquid Stocks: Evidence From China. Working Paper, Yale University.

Cochrane, J. and J. Saa-Requejo (2001). Good-Deal Asset Pricing Bounds in Incomplete Markets. *JPE 108*, 79–119.

Commissioner of Internal Revenue (2001). Respondent's Memorandum of Expert Report Differences' Bank One Corporation (Petitioner) versus Commissioner of Internal Revenue (Respondent). Docket Numbers 5759-95 and 5956-97, Judge David Laro, United States Tax Court.

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

Coval, J. and E. Stafford (2005). Asset Fire Sales (and Purchases) in Equity Markets. Working Paper, Harvard Business School.

Dai, Q. and K. Rydqvist (2003). How do Buyers and Sellers Divide The Surplus. Unpublished working paper. Binghamton University.

D'Avolio, G. (2002). The Market for Borrowing Stock. *Journal of Financial Economics 66*, 271–306.

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

Duffie, D. (1996). Special Repo Rates. *Journal of Finance 51*, 493–526.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2002). Securties Lending, Shorting, and Pricing. *Journal of Financial Economics 66*, 307–339.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005). Over-the-Counter Markets. *Econometrica 73*, 1815–1847.

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.

Edwards, A. K., L. E. Harris, and M. S. Piwowar (2004). Corporate Bond Market Transparency and Transaction Costs. Working paper, The Securities and Exchange Commission.

Fleming, M. J. and K. D. Garbade (2003). The Specials Market for U.S. Treasury Securities and the Federal Reserve's Lending Program. Unpublished working paper. Federal Reserve Bank of New York.

Froot, K. and P. O'Connell (1999). The Pricing of US Catastrophe Reinsurance. In K. Froot (Ed.), *The Financing of Catastrophe Risk*, pp. 195–232. University of Chicago Press.

Geczy, C. C., D. K. Musto, and A. V. Reed (2002). Stocks are Special Too: An Analysis of the Equity Lending Market. *Journal of Financial Economics 66*, 241–269.

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

Giroux, G. (2005). Markets of a Large Number of Interacting Agents. Working Paper, Department of Mathematics, Université de Sherbrooke, Québec.

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.

Graveline, J. and M. McBrady (2005). Who Makes On-The-Run Treasuries Special. Unpublished working paper. Stanford University.

Greenwood, R. (2005). Short and Long Term Demand Curves for Stocks: Theory and Evidence. *Journal of Financial Economics 75*, 607–650.

Grossman, S. and M. Miller (1988). Liquidity and Market Structure. *Journal of Finance 43*, 617–633.

Harris, L. and E. Gurel (1986). Price and Volume Effects Associated with Changes in the S&P 500: New Evidence of Price Pressures. *Journal of Finance 41*, 815–830.

Hradsky, G. T. and R. D. Long (1989). High-Yield Default Losses and the Return Performance of Bankrupt Debt. *Financial Analysts Journal* (July-August 1989), 38–49.

Huang, M. (2003). Liquidity Shocks and Equilibrium Liquidity Premia. *Journal of Economic Theory 109*, 104–129.

Jones, C. M. and O. A. Lamont (2002). Short Sales Constraints and Stock Returns. *Journal of Financial Economics 66*, 207–239.

Jordan, B. and S. Jordan (1997). Special Repo Rates: An Empirical Analysis. *Journal of Finance 52*, 2051–2072.

Kaul, A., V. Mehrotra, and R. Morck (2000). Demand Curces for Stocks do Slope Down: New Evidence from an Index Weights Adjustment. *Journal of Finance 55*, 893–912.

Krainer, J. and S. LeRoy (2002). Equilibrium Valuation of Illiquid Assets. *Economic Theory 19*, 223–242.

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

Lagos, R. (2005). Asset Prices and Liquidity in an Exchange Economy. Working Paper, New York University.

Lamont, O. A. and R. H. Thaler (2003). Can the Market Add and Subtract? Mispricing in Tech Stock Carve-Outs. *Journal of Political Economy 111*, 227–268.

Longstaff, F. (2002). The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices. Working paper. UCLA, forthcoming, *Journal of Business.*

Longstaff, F. (2004). Financial Claustrophobia: Asset Pricing in Illiquid Markets. Working paper. UCLA.

Longstaff, F., S. Mithal, and E. Neis (2004). Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit-Default Swap Market. *Journal of Finance forthcoming.*

Miao, J. (2004). A Search Model of Centralized and Decentralized Trade. Unpublished working paper. Department of Economics, Boston University.

Mikkelson, W. H. and M. M. Partch (1985). Stock Price Effects and Costs of Secondary Distributions. *Journal of Financial Economics 14*, 165–194.

Mitchell, M., T. Pulvino, and E. Stafford (2002). Limited Arbitrage in Equity Markets. *Journal of Finance 57*, 551–584.

Moresi, S. (1991). Three Essays in Economic Theory. Ph.D. Thesis, MIT.

Newman, Y. and M. Rierson (2003). Illiquidity Spillovers: Theory and Evidence from European Telecom Bond Issuance. Working Paper, Graduate School of Business, Stanford University.

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

Rubinstein, A. (1982). Perfect Equilibrium in a Bargaining Model. *Econometrica 50*, 97–109.

Rubinstein, A. and A. Wolinsky (1985). Equilibrium in a Market with Sequential Bargaining. *Econometrica 53*, 1133–1150.

Rubinstein, A. and A. Wolinsky (1987). Middlemen. *Quarterly Journal of Economics 102*, 581–594.

Shleifer, A. (1986). Do Demand Curves for Stocks Slope Down? *Journal of Finance 41*, 579–590.

Shleifer, A. and R. W. Vishny (1997). Limits of Arbitrage. *Journal of Finance 52*, 35–55.

Silber, W. L. (1991). Discounts on Restricted Stock: The Impact of Illiquidity on Stock Prices. *Financial Analysts Journal July-August*, 60–64.

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.

Vayanos, D. and P.-O. Weill (2005). A Search-Based Theory of The On-The-Run Phenomenon. Working Paper, London School of Economics.

Wang, J. (1994). A Model of Competitive Stock Trading Volume. *JPE 102*, 127–168.

Wang, N. (2002). Consumption and Conditionally Heteroskedastic Income. Working paper, University of Rochester.

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

Weill, P.-O. (2003). Leaning Against the Wind. Working Paper, Stanford University.

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