Foundations and Trends® in
Finance
Vol. 6, No. 4 (2011) 221–317

DOI: 10.1561/0500000014

Theories of Liquidity

By Dimitri Vayanos and Jiang Wang

Contents

Foundations and Trends® in
Finance
Vol. 6, No. 4 (2011) 221–317

DOI: 10.1561/0500000014

Theories of Liquidity

Dimitri Vayanos[1] and Jiang Wang[2]

[1] *London School of Economics, CEPR and NBER, UK, d.vayanos@lse.ac.uk*
[2] *Massachusetts Institute of Technology, CAFR and NBER, USA, wangj@mit.edu*

Abstract

We survey the theoretical literature on market liquidity. The literature traces illiquidity, i.e., the lack of liquidity, to underlying market imperfections. We consider six main imperfections: participation costs, transaction costs, asymmetric information, imperfect competition, funding constraints, and search. We address three questions in the context of each imperfection: (a) how to measure illiquidity, (b) how illiquidity relates to underlying market imperfections and other asset characteristics, and (c) how illiquidity affects expected asset returns. We nest all six imperfections within a common, unified model, and use that model to organize the literature.

1

Introduction

Under the standard Arrow–Debreu paradigm, trading in financial markets involves no frictions and liquidity is perfect. In practice, however, frictions of varying importance are present in all markets and reduce liquidity. A large and growing theoretical literature traces illiquidity, i.e., the lack of liquidity, to underlying market imperfections such as asymmetric information, different forms of trading costs, and funding constraints. It also studies how imperfections affect expected asset returns through their influence on liquidity. This literature is complemented by a large and growing empirical literature that estimates measures of illiquidity and relates them to asset characteristics and asset returns.

In this paper, we survey the theoretical literature on market liquidity. We focus on six main imperfections studied in the literature: participation costs, transaction costs, asymmetric information, imperfect competition, funding constraints, and search. These imperfections map into six different theories of illiquidity. We address three basic questions in the context of each imperfection: (a) how to measure illiquidity, (b) how illiquidity relates to underlying market imperfections and other asset characteristics, and (c) how illiquidity affects expected asset returns.

The theoretical literature on market liquidity often employs different modeling assumptions when studying different imperfections. For example, papers on trading costs typically assume life-cycle or risk-sharing motives to trade, while papers on asymmetric information often rely on noise traders. Some papers on asymmetric information further assume risk-neutral market makers who can take unlimited positions, while papers on other imperfections typically assume risk aversion or position limits. Instead of surveying this literature in a descriptive manner, we use a common, unified model to study all six imperfections that we consider, and for each imperfection we address the three basic questions within that model. Our model generates many of the key results shown in the literature, and serves as a point of reference for surveying other results derived in different or more complicated settings. We use the same model in Vayanos and Wang (2012b), where we survey both the theoretical and the empirical literature on market liquidity. This paper focuses on the theoretical literature only, surveys it more extensively, and analyzes the model in greater depth.

Our model has three periods, $t = 0, 1, 2$. In Periods 0 and 1, risk-averse agents can trade a riskless and a risky asset that pay off in Period 2. In Period 0, agents are identical so no trade occurs. In Period 1, agents can be one of two types. Liquidity demanders receive an endowment correlated with the risky asset's payoff, and need to trade to share risk. They can trade with liquidity suppliers, who receive no endowment. Agents learn whether or not they will receive the endowment in an interim period $t = 1/2$. While we model heterogeneity through endowments, our analysis would be similar for other types of heterogeneity, e.g., different beliefs or investment opportunities. Market imperfections concern trade in Period 1. We consider six imperfections, studied extensively in the theoretical literature:

1. *Participation costs*: In the perfect-market benchmark, all agents are present in the market in all periods. Thus, a seller, for example, can have immediate access to the entire population of buyers. In practice, however, agents face costs of market participation, e.g., to monitor market movements and have ready access to a financial exchange. To model costly

participation, we assume that agents must incur a cost to trade in Period 1. Consistent with the notion that participation is an ex-ante decision, we assume that agents must decide whether or not to incur the cost in Period 1/2, i.e., after learning whether or not they will receive an endowment but before observing the price in Period 1. A related imperfection is that of entry costs, e.g., learning about an asset. The cost would then concern buying the asset in Period 0.

2. *Transaction costs*: In addition to costs of market participation, agents typically pay costs when executing transactions. Transaction costs drive a wedge between the buying and selling price of an asset. They come in many types, e.g., brokerage commissions, exchange fees, transaction taxes, bid-ask spreads, and price impact. Some types of transaction costs, such as price impact, can be viewed as a consequence of other market imperfections, while other types, such as transaction taxes, can be viewed as more primitive. We assume that transaction costs concern trade in Period 1. The difference with participation costs is that the decision whether or not to incur the transaction costs is contingent on the price in Period 1.
3. *Asymmetric information*: In the perfect-market benchmark, all agents have the same information about the payoff of the risky asset. In practice, however, agents can have different information because they have access to different sources of information or have different abilities to process information from the same source. To model asymmetric information, we assume that some agents observe in Period 1 a private signal about the asset payoff. We assume that these agents are the liquidity demanders. This assumption is without loss of generality in our model. It allows us to determine how the supply of liquidity is influenced by the concern of liquidity suppliers about trading against better-informed agents.
4. *Imperfect competition*: In the perfect-market benchmark, agents are competitive and have no effect on prices. In many markets, however, some agents are large relative to others

in the sense that they can influence prices, either because of their size or because of their information advantage. We model imperfect competition by assuming that some agents can exert market power in Period 1. We mainly focus on the case where liquidity demanders behave as a single monopolist, and consider, more briefly, monopolistic behavior by liquidity suppliers. We consider both the cases where liquidity demanders have no private information on asset payoffs, and so information is symmetric, and where they observe a private signal.

5. *Funding constraints*: Agents' portfolios often involve leverage, i.e., borrow cash to establish a long position in a risky asset, or borrow a risky asset to sell it short. In the perfect-market benchmark, agents can borrow freely provided that they have enough resources to repay the loan. But as the Corporate Finance literature emphasizes, various frictions can limit agents' ability to borrow and fund their positions. We derive a funding constraint by assuming that agents cannot pledge some of their future income. Because our focus is on how the funding constraint influences the supply of liquidity, we impose it on liquidity suppliers only, i.e., assume that only they are unable to pledge their income.
6. *Search*: In the perfect-market benchmark, the market is organized as a centralized exchange. Many markets, however, have a more decentralized form of organization. For example, in over-the-counter markets, investors negotiate prices bilaterally with dealers. Locating suitable counterparties in these markets can take time and involve search. To model decentralized markets, we assume that agents do not meet in a centralized exchange in Period 1, but instead must search for counterparties. When a liquidity demander meets a supplier, they bargain bilaterally over the terms of trade.

We determine how each imperfection affects measures of illiquidity in Period 1. We consider two such measures. The first is lambda, defined as the regression coefficient of the return between Periods 0 and 1 on liquidity demanders' signed volume in Period 1. This measure characterizes the price impact of volume, which has a transitory and a permanent component. The second is price reversal, defined as minus the autocovariance of returns. This measure characterizes the importance of the transitory component in price, which in our model is entirely driven by volume. Lambda and price reversal have been derived in theoretical models focusing on specific market imperfections, and have been widely used in empirical work ever since.

In addition to the effect of imperfections on illiquidity in Period 1, we determine their effect on the ex-ante expected return as of Period 0, i.e., how does the expected return that agents require to buy the risky asset in Period 0 depend on the imperfections that they anticipate to face in Period 1. Many of the effects of imperfections that we derive within our model have been derived in the literature, albeit in a less systematic and unified manner. We highlight the links with the literature, and use more generally our model to organize and survey it. Many models in the literature can be viewed as enrichments of our model in terms of, e.g., information structure, agent characteristics, and dynamics.

Deriving the effects of the imperfections in a systematic manner within a unified model delivers new insights. We show, for example, that most imperfections raise lambda, but fewer raise price reversal. Thus, lambda is a more accurate measure of the imperfections. Intuitively, lambda measures the price impact per unit trade, while price reversal concerns the impact of the entire trade. Market imperfections generally raise the price impact per unit trade, but because they also reduce trade size, the price impact of the entire trade can decrease. We show additionally that imperfections do not always raise expected returns. The literature has shown this result for some imperfections; we examine its validity across all imperfections and identify those under which it is more likely to hold.

Our survey does not cover some important issues, either because they represent open questions on which research so far has been limited, or because covering them would detract from our main focus. Nevertheless, it is important to recognize these issues, both to put our survey in perspective and to outline promising areas for future research.

A first issue concerns the horizon of liquidity effects. The market microstructure literature focuses on liquidity effects that manifest themselves over short horizons, from minutes or hours to days or weeks. At the same time, recent work on the limits of arbitrage finds that flows can affect returns even at the longer horizons used in asset-pricing analysis, e.g., months, quarters or years. We view both horizons as relevant for the purposes of our survey — provided that the price movements under consideration are temporary departures from fundamental value caused by flows. Our model can accommodate both horizons simply by changing the length of a "period." At the same time, that length is exogenous in our model and should be derived endogenously. That would require a more detailed description of market imperfections and agents' trading needs, as well as an extension of the model along the inter-temporal dimension. Such an extension would also allow for a more complete analysis of the joint dynamics of liquidity and asset returns.

A second issue concerns the interactions between market imperfections. Most of the theoretical literature considers one imperfection at a time and does not allow for interactions. Our model also does not cover interactions, except between imperfect competition and asymmetric information. Other interactions, such as between funding constraints and asymmetric information, are interesting and have received some attention in the literature.

A related but more fundamental issue concerns the underlying economic causes of the imperfections and the ways in which imperfections are linked. Following much of the literature, we treat each imperfection as primitive. Yet, some imperfections could be the consequence of other more fundamental ones. For example, some types of transaction costs, such as price impact, can be viewed as a consequence of other imperfections, such as participation costs or asymmetric information. Moreover, if participation costs are costs to monitor market information, then costly participation could be linked to asymmetric information. Asymmetric information could also underlie the contracting frictions that give rise to funding constraints. Endogenizing some market imperfections from more fundamental frictions could further streamline, clarify and deepen the study of market liquidity. In particular, various forms of informational problems could be the underlying economic cause for various forms of imperfections.

An additional imperfection implicit in our model is that agents cannot contract ex-ante on whether they are liquidity demanders or suppliers ex-post. If they could write contracts conditional on their future trading needs, then there would be no trade ex-post and the other imperfections would not matter. Understanding the origin of this additional imperfection, and of trade more generally, is important.

A fourth issue concerns the design of the market. While we consider ways in which markets deviate from the Walrasian ideal, we do not study market design in depth. The market microstructure literature studies various dimensions of market design and shows that they can affect market performance. Such dimensions include whether liquidity is supplied by dedicated market makers or an open limit-order book, whether limit orders are visible to all traders, whether transactions are disclosed to all traders after they are executed, etc. While we survey some of that work, we conduct our analysis at a more aggregate level with less market detail, so that we can derive some key effects within a tractable unified model. The downside is that our model is not well suited for very short horizons of seconds or minutes. Our model is also not well suited for addressing the benefits of different market designs.

Related to market design is the broader institutional context. A large fraction of trading activity in financial markets is generated by specialized financial institutions, and these institutions can be important suppliers or demanders of liquidity. Following much of the literature, we model instead liquidity suppliers and demanders as individuals, thus ignoring contracting frictions and other institutional complexities. (We only consider such frictions briefly in the context of funding constraints.) The liquidity shock in our model could result from institutional frictions, but only in reduced form. The importance of financial institutions in affecting asset prices is emphasized in a rapidly growing literature on the limits of arbitrage.

Finally, we do not perform any analysis of welfare or policy (even though our model could be used for that purpose as well). For example, we do not examine how imperfections affect the welfare of different agents and what policy actions could mitigate these effects. We survey, however, some papers that consider welfare and policy issues.

Our survey is related to both market microstructure and asset pricing. We emphasize fundamental market imperfections covered in the market microstructure literature, but abstract away from the level of market detail often adopted in that literature. At the same time, we study how market imperfections affect expected asset returns — an asset-pricing exercise. Surveys with greater focus on market microstructure include the book by O'Hara (1995) for the theory, the article by Hasbrouck (2007) for the empirics, and the articles by Madhavan (2000), Biais et al. (2005), and Parlour and Seppi (2008) for both theory and empirics. Amihud et al. (2005) survey theoretical and empirical work on market liquidity and asset-pricing effects. They mainly focus on transaction costs and not on other market imperfections. We consider instead six imperfections including transaction costs, both in this survey which focuses on the theory and in Vayanos and Wang (2012b) which also surveys empirical work. Gromb and Vayanos (2010) survey the theoretical literature on the limits of arbitrage.

2

Model

There are three periods, $t = 0, 1, 2$. The financial market consists of a riskless and a risky asset that pay off in Period 2. The riskless asset is in supply of B shares and pays off one unit of a consumption good per share with certainty. The risky asset is in supply of $\bar{\theta}$ shares and pays off D units per share, where D has mean $\bar{D}$ and variance σ^2. Using the riskless asset as the numeraire, we denote by S_t the risky asset's price in Period t, where $S_2 = D$.

There is a measure one of agents, who derive utility from consumption in Period 2. Utility is exponential,

$$-\exp(-\alpha C_2), \tag{2.1}$$

where C_2 is consumption in Period 2, and $\alpha > 0$ is the coefficient of absolute risk aversion. We denote agents' wealth in Period t by W_t. Wealth in Period 2 is equal to consumption, i.e., $W_2 = C_2$. Agents are identical in Period 0, and are endowed with the per capita supply of the riskless and the risky asset. They become heterogeneous in Period 1, and this generates trade. Because all agents have the same exponential utility, there is no preference heterogeneity. We instead introduce heterogeneity through agents' endowments.

A fraction π of agents receive an endowment of the consumption good in Period 2, and the remaining fraction $1 - \pi$ receive no endowment. The endowment is $z(D - \bar{D})$ per receiving agent, where z has mean zero and variance σ_z^2 and is independent of D. Since the endowment is correlated with D, it generates a hedging demand. When, for example, $z > 0$, the endowment exposes agents to the risk that D will be low, and agents hedge against that risk by selling the risky asset. Agents learn whether or not they will receive the endowment in an interim period $t = 1/2$, and those who will receive the endowment observe z in Period 1. Thus, agents learn whether or not they will need to trade before learning the exact size of their desired trade. We assume that the endowment is perfectly correlated with D for simplicity; what matters for our analysis is that the correlation is non-zero. While we model heterogeneity through endowments, our analysis would be similar for other types of heterogeneity, e.g., different beliefs or investment opportunities.

For tractability, we assume that D and z are normal. Under normality, the endowment $z(D - \bar{D})$ can take large negative values, and this can generate an infinitely negative expected utility. To guarantee that utility is finite, we assume that the variances of D and z satisfy the condition

$$\alpha^2\sigma^2\sigma_z^2 < 1. \tag{2.2}$$

In equilibrium, agents receiving an endowment initiate trades with others to share risk. Because the agents initiating trades can be thought of as consuming market liquidity, we refer to them as liquidity demanders and denote them by the subscript d. Moreover, we refer to z as the liquidity shock. The agents who receive no endowment accommodate the trades of liquidity demanders, thus supplying liquidity. We refer to them as liquidity suppliers and denote them by the subscript s.

Because liquidity suppliers require compensation to absorb risk, the trades of liquidity demanders affect prices. Therefore, the price in Period 1 is influenced not only by the asset payoff, but also by the liquidity demanders' trades. Our measures of liquidity, defined in Section 3, are based on the price impact of these trades.

The assumptions introduced so far describe our model's perfect-market benchmark, to which we subsequently add market imperfections. We maintain the perfect-market assumption in Period 0 when determining the ex-ante effect of the imperfections, i.e., how does the expected return that agents require to buy the risky asset in Period 0 depend on the imperfections that they anticipate to face in Period 1. Imperfections in Period 0 are, in fact, not relevant in our model because agents are identical in that period and there is no trade.

We can give two interpretations to our model. Under the first interpretation, the set of agents in the model is the entire set of households in an economy. The only liquidity shocks that can then have non-trivial price impact are those large enough to be comparable to the size of the economy. Under the second interpretation, the set of agents in the model is the subset of households who participate in a specific market. Liquidity shocks can then have non-trivial price impact even when they are small relative to the size of the economy; all that is needed is that they are comparable to the size of the set of households participating in that market. That set can be smaller than the entire set of households in the economy because of participation costs. While we consider participation costs as a market imperfection (Section 4), they can be viewed as implicit in the perfect-market benchmark under the second interpretation.

An additional imperfection implicit in the perfect-market benchmark is that agents cannot write contracts in Period 0 contingent on whether they are a liquidity demander or supplier in Period 1. Thus, the market in Period 0 is incomplete in the Arrow–Debreu sense. If agents could write complete contracts in Period 0, they would not need to trade in Period 1, in which case liquidity would not matter. Complete contracts are infeasible in our model because whether an agent is a liquidity demander or supplier is private information.

3

Perfect-Market Benchmark

In this section we solve the basic model described in Section 2, assuming no market imperfections. We first compute the equilibrium, going backwards from Period 1 to Period 0. We next construct measures of illiquidity in Period 1, and study how these measures as well as the expected return of the risky asset as of Period 0 depend on the parameters of the model. Detailed derivations and proofs of the results in this and subsequent sections are in Vayanos and Wang (2010, 2012a).

3.1 Equilibrium

In Period 1, a liquidity demander chooses holdings θ_1^d of the risky asset to maximize the expected utility (2.1). Consumption in Period 2 is

$$C_2^d = W_1 + \theta_1^d(D - S_1) + z(D - \bar{D}),$$

i.e., wealth in Period 1, plus capital gains from the risky asset, plus the endowment. Therefore, expected utility is

$$-\mathrm{E}\exp\left\{-\alpha\left[W_1 + \theta_1^d(D - S_1) + z(D - \bar{D})\right]\right\}, \tag{3.1}$$

where the expectation is over D. Because D is normal, the expectation is equal to

$$-\exp\left\{-\alpha\left[W_1+\theta_1^d(\bar{D}-S_1)-\frac{1}{2}\alpha\sigma^2(\theta_1^d+z)^2\right]\right\}. \tag{3.2}$$

A liquidity supplier chooses holdings θ_1^s of the risky asset to maximize the expected utility

$$-\exp\left\{-\alpha\left[W_1+\theta_1^s(\bar{D}-S_1)-\frac{1}{2}\alpha\sigma^2(\theta_1^s)^2\right]\right\}, \tag{3.3}$$

which can be derived from (3.2) by setting $z=0$. The solution to the optimization problems is straightforward and summarized in Proposition 3.1.

Proposition 3.1. Agents' demand functions for the risky asset in Period 1 are

$$\theta_1^s = \frac{\bar{D}-S_1}{\alpha\sigma^2}, \tag{3.4a}$$

$$\theta_1^d = \frac{\bar{D}-S_1}{\alpha\sigma^2} - z. \tag{3.4b}$$

Liquidity suppliers are willing to buy the risky asset as long as its price S_1 in Period 1 is below the expected payoff $\bar{D}$, and are willing to sell otherwise. Liquidity demanders have a similar price-elastic demand function, but are influenced by the liquidity shock z. When, for example, z is positive, liquidity demanders are willing to sell because their endowment is positively correlated with the asset.

Market clearing requires that the aggregate demand equals the asset supply $\bar{\theta}$:

$$(1-\pi)\theta_1^s + \pi\theta_1^d = \bar{\theta}. \tag{3.5}$$

Substituting (3.4a) and (3.4b) into (3.5), we find

$$S_1 = \bar{D} - \alpha\sigma^2\left(\bar{\theta}+\pi z\right). \tag{3.6}$$

The price S_1 decreases in the liquidity shock z. When, for example, z is positive, liquidity demanders are willing to sell, and the price must drop so that the risk-averse liquidity suppliers are willing to buy.

In Period 0, all agents are identical. An agent choosing holdings θ_0 of the risky asset has wealth

$$W_1 = W_0 + \theta_0(S_1 - S_0) \tag{3.7}$$

in Period 1. The agent can be a liquidity supplier in Period 1 with probability $1 - \pi$, or liquidity demander with probability π. Substituting θ_1^s from (3.4a), S_1 from (3.6), and W_1 from (3.7), we can write the expected utility (3.3) of a liquidity supplier in Period 1 as

$$-\exp\left\{-\alpha\left[W_0 + \theta_0(\bar{D} - S_0) - \alpha\sigma^2\theta_0(\bar{\theta} + \pi z) + \frac{1}{2}\alpha\sigma^2(\bar{\theta} + \pi z)^2\right]\right\}. \tag{3.8}$$

The expected utility depends on the liquidity shock z since z affects the price S_1. We denote by U^s the expectation of (3.8) over z, and by U^d the analogous expectation for a liquidity demander. These expectations are agents' interim utilities in Period 1/2. An agent's expected utility in Period 0 is

$$U \equiv (1 - \pi)U^s + \pi U^d. \tag{3.9}$$

Agents choose θ_0 to maximize U. The solution to this maximization problem coincides with the aggregate demand in Period 0, since all agents are identical in that period and are in measure one. In equilibrium, aggregate demand has to equal the asset supply $\bar{\theta}$, and this determines the price S_0 in Period 0.

Proposition 3.2. The price in Period 0 is

$$S_0 = \bar{D} - \alpha\sigma^2\bar{\theta} - \frac{\pi M}{1 - \pi + \pi M}\Delta_1\bar{\theta}, \tag{3.10}$$

where

$$M = \exp\left(\tfrac{1}{2}\alpha\Delta_2\bar{\theta}^2\right)\sqrt{\frac{1 + \Delta_0\pi^2}{1 + \Delta_0(1 - \pi)^2 - \alpha^2\sigma^2\sigma_z^2}}, \tag{3.11}$$

$$\Delta_0 = \alpha^2\sigma^2\sigma_z^2, \tag{3.12a}$$

$$\Delta_1 = \frac{\alpha\sigma^2\Delta_0\pi}{1 + \Delta_0(1 - \pi)^2 - \alpha^2\sigma^2\sigma_z^2}, \tag{3.12b}$$

$$\Delta_2 = \frac{\alpha\sigma^2\Delta_0}{1 + \Delta_0(1 - \pi)^2 - \alpha^2\sigma^2\sigma_z^2}. \tag{3.12c}$$

The first term in (3.10) is the asset's expected payoff in Period 2, the second term is a discount arising because the payoff is risky, and the third term is a discount due to illiquidity (i.e., low liquidity). The risk discount is independent of the parameter σ_z^2 that measures the size of liquidity shocks, and is non-zero even when there are no shocks, i.e., $\sigma_z^2 = 0$. The illiquidity discount is instead increasing in σ_z^2, and is zero when $\sigma_z^2 = 0$. In the next section we explain why the illiquidity discount arises.

3.2 Illiquidity and its Effect on Price

We construct two measures of illiquidity in Period 1, both based on the price impact of the liquidity demanders' trades. The first measure, to which we refer as lambda or price impact, is the coefficient of a regression of the price change $S_1 - S_0$ between Periods 0 and 1 on the signed volume $\pi(\theta_1^d - \bar{\theta})$ of liquidity demanders in Period 1:

$$\lambda \equiv \frac{\mathrm{Cov}\left[S_1 - S_0, \pi(\theta_1^d - \bar{\theta})\right]}{\mathrm{Var}\left[\pi(\theta_1^d - \bar{\theta})\right]}. \tag{3.13}$$

Intuitively, when λ is large, trades have large price impact and the market is illiquid. Equation (3.6) implies that the price change between Periods 0 and 1 is

$$S_1 - S_0 = \bar{D} - \alpha\sigma^2\left(\bar{\theta} + \pi z\right) - S_0. \tag{3.14}$$

Equations (3.4b) and (3.6) imply that the signed volume of liquidity demanders is

$$\pi(\theta_1^d - \bar{\theta}) = -\pi(1-\pi)z. \tag{3.15}$$

Equations (3.13)–(3.15) imply that

$$\lambda = \frac{\alpha\sigma^2}{1-\pi}. \tag{3.16}$$

Price impact λ is higher when agents are more risk-averse (α large), the asset is riskier (σ^2 large), or liquidity suppliers are less numerous ($1-\pi$ small).

Since the signed volume of liquidity demanders is minus that of liquidity suppliers, λ is also minus the regression coefficient of the price change between Periods 0 and 1 on suppliers' signed volume in Period 1:

$$\lambda = -\frac{\mathrm{Cov}\left[S_1 - S_0, (1-\pi)(\theta_1^s - \bar{\theta})\right]}{\mathrm{Var}\left[(1-\pi)(\theta_1^s - \bar{\theta})\right]}. \tag{3.17}$$

The supplier-based definition of λ can be easier to implement empirically than the equivalent demander-based definition. Indeed, an important class of liquidity suppliers in some markets are designated market makers, and information on their trades is often available.

The second measure of illiquidity is based on the autocovariance of price changes. The liquidity demanders' trades in Period 1 cause the price to deviate from fundamental value, while the two coincide in Period 2. Therefore, price changes exhibit negative autocovariance, and more so when trades have large price impact. We use minus autocovariance

$$\gamma \equiv -\mathrm{Cov}\left(S_2 - S_1, S_1 - S_0\right), \tag{3.18}$$

as a measure of illiquidity, and refer to it as price reversal. Equations (3.6), (3.14), (3.18) and $S_2 = D$ imply that

$$\begin{aligned}\gamma &= -\mathrm{Cov}\left[D - \bar{D} + \alpha\sigma^2\left(\bar{\theta} + \pi z\right), \bar{D} - \alpha\sigma^2\left(\bar{\theta} + \pi z\right) - S_0\right] \\ &= \alpha^2\sigma^4\sigma_z^2\pi^2. \end{aligned} \tag{3.19}$$

Price reversal γ is higher when agents are more risk-averse, the asset is riskier, liquidity demanders are more numerous (π large), or liquidity shocks are larger (σ_z^2 large).[1]

The measures λ and γ have been defined in models focusing on specific market imperfections, and have been widely used in empirical work ever since. Using our model, we can examine the behavior of these measures across a variety of imperfections, and provide a broader perspective on their properties. We emphasize basic properties below, leaving more detailed results to subsequent sections.

[1] The comparative statics of autocorrelation are similar to those of autocovariance. We use autocovariance rather than autocorrelation because normalizing by variance adds unnecessary complexity.

Kyle (1985) defines λ in a model where an informed insider trades with uninformed market makers and noise traders. The price impact measured by λ concerns the aggregate order that market makers receive, which is driven both by the insider's private information and by noise trading. Our definition of λ parallels Kyle's since the trades of our liquidity demanders can be motivated by hedging or information. In Kyle, however, market makers are risk neutral, and trades affect prices only because they can contain information. Thus, λ reflects purely the amount of information that trades convey, and is permanent because the risk-neutral market makers set the price equal to their expectation of fundamental value. In general, as in our model, λ has both a transitory and a permanent component. The transitory component, present even in our perfect-market benchmark, arises because liquidity suppliers are risk averse and require a price movement away from fundamental value to absorb a liquidity shock. The permanent component arises only when information is asymmetric, for the same reasons as in Kyle.[2]

Roll (1984) links γ to the bid-ask spread, in a model where market orders cause the price to bounce between the bid and the ask. Grossman and Miller (1988) link γ to the price impact of liquidity shocks, in a model where risk-averse liquidity suppliers must incur a cost to participate in the market. In both models, price impact is purely transitory because information is symmetric. In our model, price impact has both a transitory and a permanent component, and γ isolates the effects of the transitory component. Note that besides being a measure of imperfections, γ provides a useful characterization of price dynamics: it measures the importance of the transitory component in price arising from temporary liquidity shocks, relative to the random-walk component arising from fundamentals.

Illiquidity in Period 1 lowers the price in Period 0 through the illiquidity discount, which is the third term in (3.10). To explain why the discount arises, consider the extreme case where trade in Period 1 is

[2] An alternative definition of λ, which isolates the permanent component, involves the price change between Periods 0 and 2 rather than between Periods 0 and 1. This is because the transitory deviation between price and fundamental value in Period 1 disappears in Period 2.

not allowed. In Period 0, agents know that with probability π they will receive an endowment in Period 2. The endowment amounts to a risky position in Period 1, the size of which is uncertain because it depends on z. Uncertainty about position size is costly to risk-averse agents. Moreover, the effect is stronger when agents carry a large position from Period 0 because the cost of holding a position in Period 1 is convex in the overall size of the position. (The cost is the quadratic term in (3.2) and (3.3).) Therefore, uncertainty about z reduces agents' willingness to buy the asset in Period 0.

The intuition is similar when agents can trade in Period 1. Indeed, in the extreme case where trade is not allowed, the shadow price faced by liquidity demanders moves in response to z to the point where these agents are not willing to trade. When trade is allowed, the price movement is smaller, but non-zero. Therefore, uncertainty about z still reduces agents' willingness to buy the asset in Period 0. Moreover, the effect is weaker when trade is allowed in Period 1 than when it is not (this follows from the more general result of Proposition 6.6), and therefore corresponds to a discount driven by illiquidity. Because market imperfections hinder trade in Period 1, they tend to raise the illiquidity discount in Period 0.

The illiquidity discount is the product of two terms. The first term, $\frac{\pi M}{1-\pi+\pi M}$, can be interpreted as the risk-neutral probability of being a liquidity demander: π is the true probability, and M is the ratio of marginal utilities of wealth of demanders and suppliers, where utilities are interim in Period 1/2. The second term, $\Delta_1\bar{\theta}$, is the discount that an agent would require conditional on being a demander. Since the illiquidity discount lowers the asset price in Period 0, it raises the expected return

$$E(R) = \bar{D} - S_0$$

from buying the asset in Period 0 and holding it until it pays off in Period 2. From now on, we refer to $E(R)$ simply as the asset's expected return.

The illiquidity discount is higher when liquidity shocks are larger (σ_z^2 large) and occur with higher probability (π large). It is also higher when agents are more risk averse (α large), the asset is riskier (σ^2 large), and in larger supply ($\bar{\theta}$ large). In all cases, the risk-neutral probability of being a liquidity demander is higher, and so is the discount that an agent would require conditional on being a demander. For example, an increase in any of $(\sigma_z^2, \pi, \alpha, \sigma^2)$ increases the discount required by a demander because the liquidity shock z generates higher price volatility in Period 1 (as can be seen from (3.6)). Furthermore, in the case of $(\sigma_z^2, \alpha, \sigma^2)$, the risk-neutral probability of being a demander increases because so does the ratio M of marginal utilities of wealth of demanders and suppliers: suppliers, who benefit from the higher price volatility in Period 1, become better off relative to demanders, who are hurt by this volatility. In the case of π, both M and the physical probability of being a demander increase.[3]

[3] The comparative statics of the illiquidity discount extend to its ratio relative to the discount $\alpha\sigma^2\bar{\theta}$ driven by payoff risk. Thus, while risk aversion α, payoff risk σ^2, or asset supply $\bar{\theta}$ raise the risk discount, they have an even stronger impact on the illiquidity discount. For example, an increase in α raises the risk discount because agents become more averse to payoff risk. The effect on the illiquidity discount is even stronger because not only agents become more averse to the risk of receiving a liquidity shock, but also the shock has larger price impact and hence generates more risk.

4

Participation Costs

In the perfect-market benchmark, all agents are present in the market in all periods. Thus, a seller, for example, can have immediate access to the entire population of buyers. In practice, however, agents face costs of market participation. Such costs include buying trading infrastructure or membership of a financial exchange, having capital available on short notice, monitoring market movements, etc. To model costly participation, we assume that agents must incur a cost c to trade in Period 1. Consistent with the notion that participation is an ex-ante decision, we assume that agents must decide whether or not to incur c in Period 1/2, after learning whether or not they will receive an endowment but before observing the price in Period 1. (The price depends on the liquidity shock, which is observed only in Period 1.) If the decision can be made contingent on the price in Period 1, then c is a fixed transaction cost rather than a participation cost. We consider transaction costs as a separate market imperfection (Section 5).

We structure this section as follows: first compute the equilibrium, then examine how participation costs affect the illiquidity measures and the expected return, and finally survey the literature on participation costs. We adopt the same structure in the subsequent sections, which analyze the remaining five imperfections.

4.1 Equilibrium

The price in Period 1 is determined by the participating agents. We look for an equilibrium where all liquidity demanders participate, but only a fraction $\mu > 0$ of liquidity suppliers do. Such an equilibrium exists under conditions on the parameters of the model that we derive in Corollary 4.3. We focus on the case where these conditions are met since we are interested in examining how participation costs affect the supply of liquidity. Participation costs have an effect only when the fraction of participating liquidity suppliers is interior.

Market clearing requires that the aggregate demand of participating agents equals the asset supply held by these agents. Since in equilibrium agents enter Period 1 holding $\bar{\theta}$ shares of the risky asset, market clearing takes the form

$$(1-\pi)\mu\theta_1^s + \pi\theta_1^d = [(1-\pi)\mu + \pi]\,\bar{\theta}. \tag{4.1}$$

Agents' demand functions are as in Section 3. Substituting (3.4a) and (3.4b) into (4.1), we find that the price in Period 1 is

$$S_1 = \bar{D} - \alpha\sigma^2\left[\bar{\theta} + \frac{\pi}{(1-\pi)\mu + \pi}z\right]. \tag{4.2}$$

We next determine the measure μ of participating liquidity suppliers, assuming that all liquidity demanders participate. If a supplier participates, he submits the demand function (3.4a) in Period 1. Since participation entails a cost c, wealth in Period 1 is

$$W_1 = W_0 + \theta_0(S_1 - S_0) - c. \tag{4.3}$$

Using (3.4a), (4.2) and (4.3), we can compute the interim utility U^s of a participating supplier in Period 1/2. If the supplier does not participate, holdings in Period 1 are the same as in Period 0 ($\theta_1^s = \theta_0$), and wealth in Period 1 is given by (3.7). We denote by U^{sn} the interim utility of a non-participating supplier in Period 1/2.

The participation decision is derived by comparing U^s to U^{sn} for the equilibrium choice of θ_0, which is $\bar{\theta}$. If the participation cost c is below a threshold $\underline{c}$, then all suppliers participate ($\mu = 1$). If c is above $\underline{c}$ and below a larger threshold $\bar{c}$, then suppliers are indifferent between participating or not ($U^s = U^{sn}$), and only some participate ($0 < \mu < 1$). Increasing c within that region reduces the fraction μ of participating suppliers, while maintaining the indifference condition. This is because with fewer participating suppliers, competition becomes less intense, enabling the remaining suppliers to cover their increased participation cost. Finally, if c is above $\bar{c}$, then no suppliers participate ($\mu = 0$).

Proposition 4.1. Suppose that all liquidity demanders participate. Then, the fraction of participating liquidity suppliers is

$$\mu = 1, \quad \text{if } c \leq \underline{c} \equiv \frac{\log\left(1 + \alpha^2\sigma^2\sigma_z^2\pi^2\right)}{2\alpha}, \tag{4.4a}$$

$$\mu = \frac{\pi}{1-\pi}\left(\frac{\alpha\sigma\sigma_z}{\sqrt{e^{2\alpha c} - 1}} - 1\right), \quad \text{if } \underline{c} < c < \bar{c} \equiv \frac{\log\left(1 + \alpha^2\sigma^2\sigma_z^2\right)}{2\alpha}, \tag{4.4b}$$

$$\mu = 0, \quad \text{if } c \geq \bar{c}. \tag{4.4c}$$

We next determine the participation decisions of liquidity demanders, taking those of liquidity suppliers as given.

Proposition 4.2. Suppose that a fraction $\mu > 0$ of liquidity suppliers participate. Then, a sufficient condition for all liquidity demanders to participate is

$$(1 - \pi)\mu \geq \pi. \tag{4.5}$$

Equation (4.5) requires that the measure π of liquidity demanders does not exceed the measure $(1 - \pi)\mu$ of participating suppliers. Intuitively, when demanders are the short side of the market, they stand to gain more from participation, and can therefore cover the participation cost (since suppliers do). Combining Propositions 4.1 and 4.2, we find:

Corollary 4.3. An equilibrium where all liquidity demanders and a fraction $\mu > 0$ of liquidity suppliers participate exists under the sufficient conditions $\pi \leq 1/2$ and $c \leq \hat{c} \equiv [\log\left(1 + \frac{1}{4}\alpha^2\sigma^2\sigma_z^2\right)]/2\alpha$.

For $\pi \leq 1/2$ and $c \leq \hat{c}$, only two equilibria exist: the one in the corollary and the trivial one where no agent participates because they do not expect others to participate. The same is true for π larger but close to $1/2$, and for c larger but close to $\hat{c}$.[1] When, however, c exceeds a threshold in $(\hat{c}, \bar{c})$, the equilibrium in the corollary ceases to exist, and no-participation becomes the unique equilibrium.

To determine the price in Period 0, we follow the same steps as in Section 3. The price takes a form similar to that in the perfect-market benchmark.

Proposition 4.4. The price in Period 0 is given by (3.10), where

$$M = \exp\left(\tfrac{1}{2}\alpha\Delta_2\bar{\theta}^2\right)\sqrt{\frac{1+\Delta_0\frac{\pi^2}{[(1-\pi)\mu+\pi]^2}}{1+\Delta_0\frac{(1-\pi)^2\mu^2}{[(1-\pi)\mu+\pi]^2}-\alpha^2\sigma^2\sigma_z^2}}, \tag{4.6}$$

$$\Delta_1 = \frac{\alpha\sigma^2\Delta_0\frac{\pi}{(1-\pi)\mu+\pi}}{1+\Delta_0\frac{(1-\pi)^2\mu^2}{[(1-\pi)\mu+\pi]^2}-\alpha^2\sigma^2\sigma_z^2}, \tag{4.7a}$$

$$\Delta_2 = \frac{\alpha\sigma^2\Delta_0}{1+\Delta_0\frac{(1-\pi)^2\mu^2}{[(1-\pi)\mu+\pi]^2}-\alpha^2\sigma^2\sigma_z^2}, \tag{4.7b}$$

and Δ_0 is given by (3.12a).

4.2 Participation Costs and Illiquidity

We next examine how participation costs impact the illiquidity measures and the expected return. Proceeding as in Section 3, we can

[1] Other equilibria are ruled out by the following argument. Prices and trading profits in Period 1 depend only the relative measures of participating suppliers and demanders. Therefore, if participation occurs, the fraction of either suppliers or demanders must (generically) equal one. If the fraction of demanders is less than one, then the fraction of suppliers must equal one. This is a contradiction for $\pi \leq 1/2$ because of (4.5). It is also a contradiction for π larger but close to $1/2$ because (4.5) is a sufficient condition: because liquidity demanders face the risk of liquidity shocks, they can benefit from participation more than suppliers even when they are the long side of the market. See Huang and Wang (2009, 2010) for a more detailed discussion of the nature of equilibrium under costly participation.

compute price impact λ and price reversal γ:

$$\lambda = \frac{\alpha\sigma^2}{(1-\pi)\mu}, \tag{4.8}$$

$$\gamma = \frac{\alpha^2\sigma^4\sigma_z^2\pi^2}{[(1-\pi)\mu+\pi]^2}. \tag{4.9}$$

Both measures are inversely related to the fraction μ of participating liquidity suppliers. Proposition 4.4 implies that the illiquidity discount is also inversely related to μ.

We derive comparative statics with respect to c for the equilibrium in Corollary 4.3, and consider only the region $c > \underline{c}$, where the measure μ of participating suppliers is less than one. This is without loss of generality: in the region $c \leq \underline{c}$, where all suppliers participate, prices are not affected by the participation cost and are as in the perfect-market benchmark. When $c > \underline{c}$, an increase in the participation cost lowers μ, and therefore raises price impact, price reversal, and the illiquidity discount. Since the illiquidity discount increases, so does the asset's expected return.

Proposition 4.5. An increase in the participation cost c raises price impact λ, price reversal γ, and the asset's expected return $E(R)$.

4.3 Literature

The idea that participation in financial markets is costly and hence limited dates back to Demsetz (1968). Demsetz (1968) studies the provision of immediacy, i.e., immediate execution of trades. He argues that supplying immediacy is costly but there is a demand for it. Because of the costs of supplying immediacy, only a subset of agents will choose to supply it, and they will be compensated from the price concessions they will earn from the demanders of immediacy. Demsetz (1968) identifies the suppliers of immediacy with market makers, and their compensation with the bid-ask spread.

A subsequent literature models price formation in the presence of market makers. In most of that literature, market makers are assumed to be the only suppliers of immediacy and to receive an exogenous flow of orders from the demanders of immediacy. The literature determines the bid-ask spreads chosen by market makers as a function of the process of order arrival, the degree of competition between market makers, and the inventory and risk aversion of market makers. Examples are Garman (1976), Stoll (1978), Amihud and Mendelson (1980), Ho and Stoll (1980, 1981, 1983), Cohen et al. (1981) and Mildenstein and Schleef (1983). Because of the focus on market makers' inventory, that literature is often referred to as the inventory literature.

Most of the inventory literature takes the market structure as exogenous, e.g., assumes an exogenous number of market makers. One exception is Stoll (1978), who endogenizes the number of market makers in the spirit of Demsetz (1968), taking the costs of supplying immediacy to be fixed costs of processing orders. Grossman and Miller (1988) perform a similar exercise, but emphasize more explicitly costs of market participation. Their setting is closely related to ours: a number of liquidity suppliers choose whether or not to participate in a market, and those choosing to participate pay a cost and can absorb an uncertain and exogenous order by liquidity demanders. The analysis of Grossman and Miller (1988) is closely related to the equilibrium in Periods 1/2 and 1 in our model. Grossman and Miller (1988) emphasize additionally that the bid-ask spread has drawbacks as a measure of liquidity, and suggest the use of price reversal instead. They show that price reversal increases in participation costs, consistent with our Proposition 4.5.

Grossman and Miller (1988) do not derive the effect of participation costs on ex-ante expected returns because they do not introduce our Period 0. They introduce, however, two periods after our Period 1: a Period 2 in which an offsetting liquidity shock arrives in the market, and a Period 3 in which the asset pays off. This captures the important idea that a liquidity shock experienced by some agents is absorbed first by a small set of market participants (the liquidity suppliers in Period 1) and then by a larger set of agents who gradually arrive in the market in response to the shock (the traders with the offsetting liquidity shock in Period 2). The idea that some agents arrive gradually into a market in response to profitable opportunities has received attention recently in the search literature reviewed in Section 9. Duffie (2010) and Suominen and Rinne (2011) model a similar effect in a centralized market. They assume that some agents arrive into a market infrequently with liquidity shocks. These shocks are absorbed by market makers present in the market in all periods, and by other infrequent traders arriving in future periods who can trade with market makers.

Huang and Wang (2009) study how participation costs affect both the demand for immediacy, which Grossman and Miller (1988) treat as exogenous, and the supply. They assume that liquidity shocks are opposite across agents and so do not affect the price in the absence of participation costs. Participation costs lower the price because sellers are more willing to participate than buyers. The intuition is that sellers receive a larger risky endowment, and are hence more concerned about the risk that an additional shock will leave them with a large risk exposure. This effect of participation costs on ex-ante expected returns is closely related to the one that we derive in Period 0 of our model. Huang and Wang (2010) employ a similar framework as in Huang and Wang (2009) to study welfare questions. They show, in particular, that the market can provide less liquidity than the social optimum.

The costs of market participation in our model concern Period 1, which is after agents have bought the asset. Costs to participate in the market in Period 0 and to buy the asset can be interpreted as entry costs, e.g., learning about the asset. Goldsmith (1976), Mayshar (1979) and Merton (1987) show that entry costs induce agents to under-invest and under-diversify, and typically reduce asset prices. Entry costs would have a similar effect in our model: they would render agents less willing to buy the asset in Period 0, and hence would lower the Period 0 price. Mankiw and Zeldes (1991) conjecture that limited investor participation in the stock market can render stocks cheaper relative to bonds, explaining the equity premium puzzle of Mehra and Prescott (1985). Basak and Cuoco (1998) show that when some investors cannot participate in the stock market, stocks' expected excess returns relative to bonds increase, and interest rates decrease.

Pagano (1989a) and Allen and Gale (1994) show that entry costs can result in multiple equilibria: high-volatility ones, where few agents enter the market for an asset, causing volatility to be high and entry to be undesirable, and low-volatility ones, where many agents enter. Key for the multiplicity in Pagano (1989a) is the feedback from asset prices to equity issuance by firms, and in Allen and Gale (1994) the heterogeneity between investors.[2]

[2] A different type of multiplicity arises when agents can choose between market venues to trade an asset. Agents prefer to trade in a venue where others are trading, and this causes concentration of trade in one venue (Pagano, 1989b). A related multiplicity result arises in our model because there exists one equilibrium in which there is market participation in Period 1 and one equilibrium in which no agent participates.

5

Transaction Costs

In addition to costs of market participation, agents typically pay costs when executing transactions. Transaction costs drive a wedge between the buying and selling price of an asset. They come in many types, e.g., brokerage commissions, exchange fees, transaction taxes, bid-ask spreads, price impact. Some types of transaction costs can be viewed as a consequence of other market imperfections: for example, Section 4 shows that costly participation can generate price-impact costs. Other types of costs, such as transaction taxes, can be viewed as more primitive. We assume that transaction costs concern trade in Period 1. The difference with the participation costs of Section 4 is that the decision whether or not to incur the transaction costs is contingent on the price in Period 1.

We focus on the case where transaction costs are proportional to transaction size, and for simplicity assume that proportionality concerns the number of shares rather than the dollar value. Denoting by κ the cost per unit of shares traded and by θ_t the number of shares that an agent holds in Period $t = 0, 1$, proportional costs take the form $\kappa\,|\theta_1 - \theta_0|$. We assume that the liquidity shock z is drawn from a general distribution that is symmetric around zero with density $f(z)$; specializing to a normal distribution does not simplify the analysis.

5.1 Equilibrium

Transaction costs generate a bid-ask spread in Period 1. An agent buying one share pays the price S_1 plus the transaction cost κ, and so faces an effective ask price $S_1 + \kappa$. Conversely, an agent selling one share receives S_1 but pays κ, and so faces an effective bid price $S_1 - \kappa$. The bid-ask spread is independent of transaction size because transaction costs are proportional. Because of the spread, trade occurs only if the liquidity shock z is sufficiently large. Suppose, for example, that $z > 0$, in which case liquidity demanders value the asset less than liquidity suppliers. If liquidity suppliers buy, their demand function is similar to that in Section 3 (Equation (3.4a)), but with $S_1 + \kappa$ taking the place of S_1, i.e.,

$$\theta_1^s = \frac{\bar{D} - S_1 - \kappa}{\alpha\sigma^2}. \tag{5.1}$$

Conversely, if liquidity demanders sell, their demand function is also similar to that in Section 3 (Equation (3.4b)), but with $S_1 - \kappa$ taking the place of S_1, i.e.,

$$\theta_1^d = \frac{\bar{D} - S_1 + \kappa}{\alpha\sigma^2} - z. \tag{5.2}$$

Since in equilibrium agents enter Period 1 holding $\bar{\theta}$ shares of the risky asset, trade occurs if there exists a price S_1 such that $\theta_1^s > \bar{\theta}$ and $\theta_1^d < \bar{\theta}$. Using (5.1) and (5.2), we can write these conditions as

$$\kappa < \bar{D} - S_1 - \alpha\sigma^2\bar{\theta} < \alpha\sigma^2 z - \kappa.$$

Therefore, trade occurs if $z > \frac{2\kappa}{\alpha\sigma^2} \equiv \hat{\kappa}$, i.e., the liquidity shock z is large relative to the transaction cost κ. The price can be determined by substituting (5.1) and (5.2) into the market-clearing equation (3.5). Repeating the analysis for $z < 0$, we can derive the following proposition.

Proposition 5.1. The equilibrium in Period 1 is as follows:

- $|z| \leq \hat{\kappa}$: Agents do not trade;
- $|z| > \hat{\kappa}$: All agents trade and the price is

$$S_1 = \bar{D} - \alpha\sigma^2 \left[\bar{\theta} + \pi z + \hat{\kappa}\left(\frac{1}{2} - \pi\right)\operatorname{sign}(z)\right]. \tag{5.3}$$

The effect of transaction costs on the price depends on the relative measures of liquidity suppliers and demanders. Suppose, for example, that $z > 0$. In the absence of transaction costs, liquidity demanders sell and the price drops. Because transaction costs deter liquidity suppliers from buying, they tend to depress the price, amplifying the effect of z. At the same time, transaction costs deter liquidity demanders from selling, and this tends to raise the price, dampening the effect of z. The overall effect depends on agents' relative measures. If $\pi < 1/2$ (more suppliers than demanders), the impact on suppliers dominates, and transaction costs amplify the effect of z. The converse holds if $\pi > 1/2$. The price in Period 0 takes a form similar to that in the perfect-market benchmark.[1]

Proposition 5.2. The price in Period 0 is given by (3.10), where

$$M = \frac{\int_0^{\hat{\kappa}} \exp\left(\frac{1}{2}\alpha^2\sigma^2 z^2\right) \mathrm{ch}(\alpha^2\sigma^2\bar{\theta}z) f(z)dz + \int_{\hat{\kappa}}^{\infty} \Gamma(z)\mathrm{ch}(\alpha^2\sigma^2\bar{\theta}z) f(z)dz}{\int_0^{\hat{\kappa}} f(z)dz + \int_{\hat{\kappa}}^{\infty} \exp\left[-\frac{1}{2}\alpha^2\sigma^2\pi^2(z-\hat{\kappa})^2\right] f(z)dz}, \tag{5.4}$$

$$\Delta_1 = \frac{\alpha\sigma^2 \left[\begin{array}{l} \int_0^{\hat{\kappa}} \exp\left(\frac{1}{2}\alpha^2\sigma^2 z^2\right) \mathrm{sh}\left(\alpha^2\sigma^2\bar{\theta}z\right) z f(z)dz \\ \qquad + \int_{\hat{\kappa}}^{\infty} \Gamma(z)\mathrm{sh}(\alpha^2\sigma^2\bar{\theta}z)[\pi z + (1-\pi)\hat{\kappa}] f(z)dz \end{array}\right]}{\bar{\theta} \left[\begin{array}{l} \int_0^{\hat{\kappa}} \exp\left(\frac{1}{2}\alpha^2\sigma^2 z^2\right) \mathrm{ch}(\alpha^2\sigma^2\bar{\theta}z) f(z)dz \\ \qquad + \int_{\hat{\kappa}}^{\infty} \Gamma(z)\mathrm{ch}(\alpha^2\sigma^2\bar{\theta}z) f(z)dz \end{array}\right]}, \tag{5.5}$$

$$\Gamma(z) = \exp\left[\frac{1}{2}\alpha^2\sigma^2 z^2 - \frac{1}{2}\alpha^2\sigma^2(1-\pi)^2(z-\hat{\kappa})^2\right]. \tag{5.6}$$

---

[1] Extending our analysis to fixed costs is more complicated because agents' optimization problems become non-convex. Non-convexity can give rise to multiple solutions, meaning that agents of the same type (suppliers or demanders) can fail to take the same action. Suppose, for example, that all agents start with the same position $\theta_0 = \bar{\theta}$ in Period 0. As with proportional costs, all agents trade in Period 1 if the liquidity shock z is large, while no agent trades if z is small. For intermediate values of z, however, some agents pay the fixed cost and trade, while others of the same type do not trade.

A further complication arising from non-convexity is that $\theta_0 = \bar{\theta}$ is not an equilibrium. Indeed, consider a deviation from $\theta_0 = \bar{\theta}$ in either direction. The trades that become profitable in the margin are those whose surplus equals the fixed cost. But while the net surplus of these trades is zero, the marginal surplus (i.e., the derivative with respect to θ_0) is non-zero. Thus, expected utility at $\theta_0 = \bar{\theta}$ has a local minimum and a kink, implying that identical agents in Period 0 choose different positions in equilibrium.

5.2 Transaction Costs and Illiquidity

We next examine how transaction costs impact the illiquidity measures and the expected return. Because transaction costs deter liquidity suppliers from trading, they raise price impact λ. Note that λ rises even when transaction costs dampen the effect of the liquidity shock z on the price. Indeed, dampening occurs not because of enhanced liquidity supply, but because liquidity demanders scale back their trades.

Proposition 5.3. Price impact λ is

$$\lambda = \frac{\alpha\sigma^2}{1-\pi}\left[1 + \frac{\hat{\kappa}}{2\pi}\frac{\int_{\hat{\kappa}}^{\infty}(z-\hat{\kappa})f(z)dz}{\int_{\hat{\kappa}}^{\infty}(z-\hat{\kappa})^2 f(z)dz}\right], \tag{5.7}$$

and is higher than without transaction costs ($\kappa = 0$).

Defining price reversal γ involves a slight complication because for small values of z there is no trade in Period 1, and therefore the price S_1 is not uniquely defined. We define price reversal conditional on trade in Period 1. The empirical counterpart of our definition is that no-trade observations are dropped from the sample. Transaction costs affect price reversal both because they limit trade to large values of z, and because they impact the price conditional on trade occuring. The first effect raises price reversal. The second effect works in the same direction when transaction costs amplify the effect of z on the price, i.e., when $\pi < 1/2$.

Proposition 5.4. Price reversal γ is

$$\gamma = \alpha^2\sigma^4\frac{\int_{\hat{\kappa}}^{\infty}\left[\pi z + \left(\frac{1}{2}-\pi\right)\hat{\kappa}\right]^2 f(z)dz}{\int_{\hat{\kappa}}^{\infty} f(z)dz}. \tag{5.8}$$

It is increasing in the transaction cost coefficient κ if $\pi \leq 1/2$.

Because transaction costs hinder trade in Period 1, a natural conjecture is that they raise the illiquidity discount. When, however, $\pi \approx 1$, transaction costs can lower the discount. The intuition is that for $\pi \approx 1$

liquidity suppliers are the short side of the market and stand to gain the most from trade. Therefore, transaction costs hurt them the most, and reduce the ratio M of marginal utilities of wealth of demanders and suppliers. This lowers the risk-neutral probability $\pi M/(1-\pi+\pi M)$ of being a demander, and can lower the discount. Transaction costs always raise the discount, and hence the asset's expected return, when $\pi \leq 1/2$.

Proposition 5.5. The asset's expected return $E(R)$ is decreasing in the transaction cost coefficient κ if $\pi \leq 1/2$.

We can sharpen the results of Propositions 5.4 and 5.5 by assuming specific distributions for the liquidity shock z. When z is drawn from a two-point distribution, transaction costs raise price reversal γ for all values of π, but lower the illiquidity discount for $\pi \approx 1$. When z is normal, transaction costs raise γ for all values of π, and numerical calculations suggest that they also raise the discount for all values of π.

5.3 Literature

Early papers on the effects of transaction costs are Amihud and Mendelson (1986) and Constantinides (1986). Constantinides (1986) derives the optimal investment policy of an infinitely lived agent, who can trade a riskless and a risky asset. The return of the riskless asset is constant over time, and that of the risky asset is *i.i.d.* The risky asset carries transaction costs, which are proportional to the dollar value traded. Because the agent has CRRA preferences, the optimal policy in the absence of transaction costs is to maintain a constant fraction of wealth invested the risky asset, as in Merton (1971). In the presence of transaction costs, the agent instead prevents this fraction from exiting an interval. When the fraction is strictly inside the interval, the agent does not trade. The agent incurs a small utility loss from transaction costs, even though he trades infinitely often in their absence. Intuitively, the derivative of the utility at the optimal policy is zero, and hence a deviation from that policy results in a second-order loss.

The solution of Constantinides (1986) is approximate because consumption is assumed to be an exogenous constant fraction of wealth. Davis and Norman (1990) provide an exact solution. Fleming et al. (1990), and Dumas and Luciano (1991) do the same in the more tractable case where the agent consumes only at the end of his investment horizon. To eliminate horizon effects, they focus on the limit where the horizon converges to infinity. Liu and Loewenstein (2002) consider explicitly the finite-horizon case. Balduzzi and Lynch (1999), Lynch and Balduzzi (2000), Liu (2004), Jang et al. (2007) and Lynch and Tan (2011) consider richer settings, involving multiple risky assets and predictable returns.

While Constantinides (1986) and the subsequent literature mainly emphasize portfolio optimization, they also explore implications for equilibrium asset prices. They do this by giving the agent a choice between two economies: one in which the risky asset carries transaction costs, and one in which it does not but its expected return is lower. They interpret the reduction in expected return that would make the agent indifferent between the two economies as an equilibrium effect of transaction costs. Whether this effect would arise in an explicit equilibrium model, such as those that we survey in the remainder of this section, is doubtful; for example, the effect should depend on the scarcity of the more liquid asset, but asset supply is not taken into consideration. This exercise, however, provides an intuitive metric to express the utility loss from transaction costs.

Amihud and Mendelson (1986) build an equilibrium model, in which agents are risk neutral and have different investment horizons. Upon entering the economy, agents can invest in a set of assets that differ in transaction costs. Agents must sell their assets when they exit the economy, and exit rates are independent of age but can differ across agents. Assets with high transaction costs trade at a lower price in equilibrium. Moreover, they are held by agents with long investment horizons, i.e., low exit rates, who can amortize the costs over a longer period. Each agent holds only one asset, the one maximizing expected return net of transaction costs amortized over the agent's horizon. The effect of transaction costs on asset prices is concave. Indeed, the price differential between one asset and its next closest in terms of transaction costs is determined by the "marginal" investor who is indifferent between the two assets. Since the marginal investor in the case of assets with high transaction costs has a long horizon and hence is less concerned about costs, the price differential between these assets is smaller than for low-cost assets.

Aiyagari and Gertler (1991) and Vayanos and Vila (1999) allow for intertemporal consumption and risk aversion in a setting with two riskless assets, one of which carries transaction costs. The latter paper shows that a decrease in the supply of the more liquid asset increases the liquidity premium, i.e., the premium investors are willing to pay to hold that asset. This is in the spirit of Amihud and Mendelson (1986): since the horizon of the marginal investor becomes shorter, the investor is more concerned about transaction costs. Heaton and Lucas (1996) assume that the asset carrying transaction costs is risky and agents trade to smooth labor income shocks. A calibration of their model reveals that transaction costs have small effects on asset prices unless agents face borrowing constraints.

Vayanos (1998) re-examines the effects of transaction costs in a setting with multiple risky assets and risk averse agents. Agents hold a diversified portfolio at all times, but when they need to reduce their risk exposure they sell disproportionately more of the assets with low transaction costs. Moreover, because transaction costs make agents less willing not only to buy but also to sell an asset, assets with high costs can trade at higher prices than assets with low costs. This result, which also holds in Period 1 of our model, cannot arise when agents are risk neutral or assets are riskless because of a "dominance" argument: since assets are perfect substitutes except for transaction costs, agents would not buy assets with high costs if these trade at higher prices than assets with low costs. Furthermore, the marginal-investor pricing derived in Amihud and Mendelson (1986) does not hold since agents hold diversified portfolios and hence are all marginal for an asset pair.

Huang (2003) assumes stochastic liquidation needs and two riskless assets, one of which carries transaction costs. He shows that transaction costs can generate a strict preference for diversification even though the assets are riskless. This is because returns net of transaction costs are risky: investing in the less liquid asset yields a low payoff if an agent needs to sell on short notice, and a high payoff otherwise.

Lo et al. (2004) assume that agents trade to share risk and have access to a riskless asset carrying no costs and a risky asset carrying fixed costs, i.e., independent of transaction size. They show that transaction costs hinder risk sharing, as in Period 1 of our model, and this causes the price of the risky asset to decrease, as in Period 0 of our model. Because agents in their model have high-frequency trading needs, small fixed costs have a strong effect on the price and the expected return of the risky asset.

More recent work on transaction costs emphasizes the time variation in these costs and in the liquidity premia per unit of the costs. Acharya and Pedersen (2005) assume that investors have a one-period horizon and transaction costs are stochastic. They show that part of the costs' price effect is through a risk premium. This is because transaction costs impact the covariance between an asset's return net of costs and the net return of the market portfolio. For example, if an asset's transaction costs increase when the costs of the market portfolio increase or when the market portfolio's dividends decrease, this adds to the asset's risk and causes the asset price to decrease. Beber et al. (2012) examine the effects of stochastic transaction costs when investors differ in their horizons.

Vayanos (2004) explores time variation in investor horizons, assuming constant transaction costs. He assumes that investors are fund managers subject to withdrawals when their performance drops below a threshold, and that the volatility of asset dividends is time-varying. During volatile times, fund managers' horizons shorten because their performance is more likely to drop below the threshold. This causes liquidity premia per unit of transaction costs to increase. It also causes the market betas of assets with high transaction costs to increase precisely during the times when the market is the most risk averse.

Papers on time varying transaction costs and liquidity premia show that the traditional CAPM should be augmented by pricing factors relating to aggregate liquidity. These factors are aggregate transaction costs in Acharya and Pedersen (2005) and Beber et al. (2012), and volatility (which correlates with liquidity premia) in Vayanos (2004).

Buss and Dumas (2011) and Buss et al. (2011) develop numerical algorithms to solve dynamic general equilibrium models with transaction costs and general model primitives. Buss et al. (2011) assume multiple risky assets and labor income shocks, and show that transaction costs have small price effects. Buss and Dumas (2011) show that deterministic transactions costs give rise to time-variation in measures of illiquidity, such as price impact and volume, and this variation can be a priced factor.

Most of the papers mentioned so far adopt specific functional forms for the primitives of the model, e.g., agents' utility functions and asset payoffs. This is because without functional restrictions it is difficult to derive closed-form solutions for portfolio optimization and equilibrium under transaction costs. (The same applies for other market imperfections that we consider in this survey.) A different set of papers derive more basic properties of markets with transaction costs without resorting to specific functional forms. For example, Jouini and Kallal (1995) show that under transaction costs, no-arbitrage does not require the discounted price to be a martingale but only that a martingale lies between the bid and ask prices. Luttmer (1996) shows that transaction costs weaken significantly the restrictions that the Euler equation of portfolio optimization imposes on the link between consumption and asset prices. Leland and Rubinstein (1985), Edirisinghe et al. (1993), and Soner et al. (1995) show that transaction costs also weaken the no-arbitrage bounds of option pricing.

6

Asymmetric Information

In the perfect-market benchmark, all agents have the same information about the payoff of the risky asset. In practice, however, agents can have different information because they have access to different sources of information or have different abilities to process information from the same source. We model asymmetric information through a private signal s about the asset payoff D that some agents observe in Period 1. The signal is

$$s = D + \epsilon \tag{6.1}$$

where ϵ is normal with mean zero and variance σ_ϵ^2, and is independent of (D, z). We assume that liquidity demanders, who observe the liquidity shock z in Period 1, are also the only ones to observe s. This assumption is without loss of generality: even if liquidity demanders do not observe the signal, they can infer it perfectly from the price because they observe the liquidity shock. Asymmetric information can therefore exist only if some liquidity suppliers are uninformed. We assume that they are all uninformed for simplicity. Note that because liquidity suppliers are uninformed, our model determines how the supply of liquidity is influenced by suppliers' concern about trading against better-informed agents.

6.1 Equilibrium

The price in Period 1 incorporates the signal of liquidity demanders, and therefore reveals information to liquidity suppliers. To solve for equilibrium, we follow the standard rational expectations equilibrium (REE) procedure to conjecture a price function, i.e., a relationship between the price and the signal, then determine how agents use their knowledge of the price function to learn about the signal and formulate demand functions, and finally confirm that the conjectured price function clears the market.

We conjecture a price function that is affine in the signal s and the liquidity shock z, i.e.,

$$S_1 = a + b(s - \bar{D} - cz) \tag{6.2}$$

for three constants (a, b, c). For expositional convenience, we set $\xi \equiv s - \bar{D} - cz$. We also refer to the price function as simply the price.

Agents use the price and their private information to form a posterior distribution about the asset payoff D. For a liquidity demander, the price conveys no additional information relative to observing the signal s. Given the joint normality of (D, ϵ), D remains normal conditional on $s = D + \epsilon$, with mean and variance

$$\mathrm{E}[D|s] = \bar{D} + \beta_s(s - \bar{D}), \tag{6.3a}$$

$$\sigma^2[D|s] = \beta_s \sigma_\epsilon^2, \tag{6.3b}$$

where $\beta_s \equiv \sigma^2/(\sigma^2 + \sigma_\epsilon^2)$. For a liquidity supplier, the only information is the price S_1, which is equivalent to observing ξ. Conditional on ξ (or S_1), D is normal with mean and variance

$$\mathrm{E}[D|S_1] = \bar{D} + \beta_\xi \xi = \bar{D} + \frac{\beta_\xi}{b}(S_1 - a), \tag{6.4a}$$

$$\sigma^2[D|S_1] = \beta_\xi(\sigma_\epsilon^2 + c^2\sigma_z^2), \tag{6.4b}$$

where $\beta_\xi \equiv \sigma^2/\sigma_\xi^2$ and $\sigma_\xi^2 \equiv \sigma^2 + \sigma_\epsilon^2 + c^2\sigma_z^2$. Agents' optimization problems are as in Section 3, with the conditional distributions of D replacing the unconditional one. Proposition 6.1 summarizes the solution to these problems.

Proposition 6.1. Agents' demand functions for the risky asset in Period 1 are

$$\theta_1^s = \frac{\mathrm{E}[D|S_1] - S_1}{\alpha\sigma^2[D|S_1]}, \tag{6.5a}$$

$$\theta_1^d = \frac{\mathrm{E}[D|s] - S_1}{\alpha\sigma^2[D|s]} - z. \tag{6.5b}$$

Substituting (6.5a) and (6.5b) into the market-clearing equation (3.5), we find

$$(1-\pi)\frac{\mathrm{E}[D|S_1] - S_1}{\alpha\sigma^2[D|S_1]} + \pi\left(\frac{\mathrm{E}[D|s] - S_1}{\alpha\sigma^2[D|s]} - z\right) = \bar{\theta}. \tag{6.6}$$

The price (6.2) clears the market if (6.6) is satisfied for all values of (s,z). Substituting S_1, $\mathrm{E}[D|s]$, and $\mathrm{E}[D|S_1]$ from (6.2), (6.3a) and (6.4a), we can write (6.6) as an affine equation in (s,z). Therefore, (6.6) is satisfied for all values of (s,z) if the coefficients of (s,z) and of the constant term are equal to zero. This yields a system of three equations in (a,b,c), solved in Proposition 6.2.

Proposition 6.2. The price in Period 1 is given by (6.2), where

$$a = \bar{D} - \alpha(1-b)\sigma^2\bar{\theta}, \tag{6.7a}$$

$$b = \frac{\pi\beta_s\sigma^2[D|S_1] + (1-\pi)\beta_\xi\sigma^2[D|s]}{\pi\sigma^2[D|S_1] + (1-\pi)\sigma^2[D|s]}, \tag{6.7b}$$

$$c = \alpha\sigma_\epsilon^2. \tag{6.7c}$$

To determine the price in Period 0, we follow the same steps as in Section 3. The calculations are more complicated because expected utilities in Period 1 are influenced by two random variables (s,z) rather than only z. The price in Period 0, however, takes the same general form as in the perfect-market benchmark.

Proposition 6.3. The price in Period 0 is given by (3.10), where M is given by (3.11),

$$\Delta_0 = \frac{(b-\beta_\xi)^2(\sigma^2+\sigma_\epsilon^2+c^2\sigma_z^2)}{\sigma^2[D|S_1]\pi^2}, \tag{6.8a}$$

$$\Delta_1 = \frac{\alpha^3 b\sigma^2(\sigma^2+\sigma_\epsilon^2)\sigma_z^2}{1+\Delta_0(1-\pi)^2-\alpha^2\sigma^2\sigma_z^2}, \tag{6.8b}$$

$$\Delta_2 = \frac{\alpha^3\sigma^4\sigma_z^2\left[1+\frac{(\beta_s-b)^2(\sigma^2+\sigma_\epsilon^2)}{\sigma^2[D|s]}\right]}{1+\Delta_0(1-\pi)^2-\alpha^2\sigma^2\sigma_z^2}. \tag{6.8c}$$

6.2 Asymmetric Information and Illiquidity

We next examine how asymmetric information impacts the illiquidity measures and the expected return. When some agents observe a private signal, this not only generates dispersion in information across agents, but also renders each agent more informed because the signal is partially revealed through the price. The improvement in each agent's information is not a distinguishing feature of asymmetric information: information can also improve if all agents observe a public signal. To focus on the dispersion in information, which is what distinguishes asymmetric information, we compare with two symmetric-information benchmarks: the no-information case, where information is symmetric because no agent observes the signal s, and the full-information case, where all agents observe s. The analysis in Section 3 concerns the no-information case, but can be extended to the full-information case. Price impact λ and price reversal γ under full information are given by (3.16) and (3.19), respectively, where σ^2 is replaced by $\sigma^2[D|s]$.

Proposition 6.4. Price impact λ under asymmetric information is

$$\lambda = \frac{\alpha\sigma^2[D|S_1]}{(1-\pi)\left(1-\frac{\beta_\xi}{b}\right)}. \tag{6.9}$$

Price impact is highest under asymmetric information and lowest under full information. Moreover, price impact under asymmetric information increases when the private signal (6.1) becomes more precise, i.e., when σ_ϵ^2 decreases.

Proposition 6.4 shows that price impact is higher under asymmetric information than under either of the two symmetric-information benchmarks. This comparison is driven by an uncertainty and a learning effect. Price impact increases in the uncertainty faced by liquidity suppliers, measured by their conditional variance of the asset payoff. Because of this uncertainty effect, price impact tends to be lowest under full information, since liquidity suppliers observe the signal perfectly, next lowest under asymmetric information, since the signal is partially revealed to liquidity suppliers through the price, and highest under no information.

An additional source of price impact, present only under asymmetric information, is that liquidity suppliers seek to learn the signal from the price. Because, for example, liquidity suppliers attribute selling pressure partly to a low signal, they require a larger price drop to buy. This learning effect corresponds to the term β_ξ/b in (6.9), which lowers the denominator and raises price impact λ. The learning effect works in the same direction as the uncertainty effect when comparing asymmetric to full information, but in the opposite direction when comparing asymmetric to no information. Proposition 6.4 shows that in the latter comparison the learning effect dominates.

While price impact is unambiguously higher under asymmetric information, the same is not true for price reversal. Indeed, consider two extreme cases. If $\pi \approx 1$, i.e., almost all agents are liquidity demanders (informed), then the price processes under asymmetric and full information approximately coincide, and so do the price reversals. Since, in addition, liquidity suppliers face more uncertainty under no information than under full information, price reversal is highest under no information. If instead $\pi \approx 0$, i.e., almost all agents are liquidity suppliers (uninformed), then price impact λ converges to infinity (order $1/\pi$) under asymmetric information. This is because the trading volume of liquidity demanders converges to zero, but the volume's informational content remains unchanged. Because of the high price impact, price reversal is highest under asymmetric information.

Proposition 6.5. Price reversal γ under asymmetric information is

$$\gamma = b(b - \beta_\xi)(\sigma^2 + \sigma_\epsilon^2 + c^2\sigma_z^2). \tag{6.10}$$

Price reversal is lowest under full information. It is highest under asymmetric information if $\pi \approx 0$, and under no information if $\pi \approx 1$.

The analysis of the illiquidity discount involves an effect that goes in the direction opposite to the uncertainty effect. This is that information revealed about the asset payoff in Period 1 reduces uncertainty and hence the scope for risk sharing. Less risk sharing, in turn, renders agents less willing to buy the asset in Period 0 and raises the illiquidity discount. The negative effect of information on risksharing and welfare has been shown in Hirshleifer (1971). We derive the implications of the Hirshleifer effect for asset pricing: Proposition 6.6 shows that the reduced scope for risksharing in Period 1 lowers the asset price in Period 0 and raises the illiquidity discount.

Because of the Hirshleifer effect, the illiquidity discount under full information is higher than under no information — a comparison which is exactly the reverse than for the measures of illiquidity. A corollary of this result is that the illiquidity discount under no trade is higher than in the perfect-market benchmark of Section 3. Indeed, the perfect-market benchmark corresponds to the no-information case, while no trade is a special case of full information when the signal (6.1) is perfectly precise ($\sigma_\epsilon^2 = 0$).[1]

The Hirshleifer effect implies that the illiquidity discount under asymmetric information should be between that under no and under full information. The discount under asymmetric information, however, is also influenced by the learning effect, which raises price impact, reduces

[1] Recall from Section 3 that the illiquidity discount is the product of $\pi M/(1 - \pi + \pi M)$, the risk-neutral probability of being a liquidity demander, times $\Delta_1\bar{\theta}$, the discount that an agent would require conditional on being a demander. No trade renders both demanders and suppliers worse off relative to the perfect-market benchmark, and hence has an ambiguous effect on the ratio M of their marginal utilities of wealth. The increase in the illiquidity discount is instead driven by the increase in the discount $\Delta_1\bar{\theta}$ required by a demander.

the scope for risk sharing and hence raises the discount. The learning effect works in the same direction as the Hirshleifer effect when comparing asymmetric to no information, but in the opposite direction when comparing asymmetric to full information. Proposition 6.6 shows that in the latter comparison the learning effect dominates. Therefore, the illiquidity discount, and hence the asset's expected return, is higher under asymmetric information than under either of the two symmetric-information benchmarks.

Proposition 6.6. The asset's expected return $E(R)$ is highest under asymmetric information and lowest under no information.

6.3 Literature

The analysis of REE with asymmetric information was pioneered by Grossman (1976). Grossman (1976) assumes that agents observe private signals about the payoff of a risky asset, which are of equal quality and independent conditional on the payoff. The equilibrium price of the risky asset reveals the average of agents' signals, which is a sufficient statistic for all the signals because of normality. Hence, the price aggregates information perfectly.

Grossman and Stiglitz (1980) assume that some agents observe a common signal about the payoff of a risky asset and the remaining agents observe no signal. Following some of the literature, we term this information structure "asymmetric information structure," and that in Grossman (1976) as "differential information structure." Grossman and Stiglitz (1980) allow additionally for the supply of the risky asset to be stochastic. With a deterministic supply, the price reveals perfectly the signal of the informed agents, and hence the uninformed can achieve the same utility as the informed. With a stochastic supply instead, the informed can achieve higher utility. The analysis of Grossman and Stiglitz (1980) is closely related to the equilibrium in Period 1 of our model, except that we introduce noise in the price through endowments rather than through the asset supply. Diamond and Verrecchia (1981) are first to use this modeling trick, and do so in a differential information model.

Grossman (1976) and Grossman and Stiglitz (1980) derive two basic paradoxes relating to information aggregation. Since the price in Grossman (1976) aggregates perfectly agents' private signals, agents should form their asset demand based only on the price and not on their signals. The paradox then is how can the price aggregate the signals. A second paradox is that if the price in Grossman and Stiglitz (1980) reveals perfectly the signal of the informed agents, then why would the informed be willing to commit resources to acquire their signal.

Both paradoxes can be resolved by introducing noise, e.g., through stochastic asset supply. Grossman and Stiglitz (1980) show that with stochastic supply, the informed can achieve higher utility than the uninformed and hence can have an incentive to acquire costly information. This has the important implication that markets cannot be fully efficient when information acquisition is costly because information will be acquired only when the price is not fully revealing.[2] Hellwig (1980) introduces stochastic supply in a differential information model, which generalizes Grossman (1976) by allowing for heterogeneity in signal quality and agent risk aversion. He shows that the price does not aggregate information perfectly, and hence agents have an incentive to use both the price and their private signal when forming their asset demand.

All papers mentioned so far assume that agents can trade one riskless and one risky asset over one period. Admati (1985) extends the analysis to multiple risky assets, while also allowing for a general correlation structure among asset payoffs, asset supplies, and agents' private signals. She shows that because signals about one asset are also informative about the payoff and supply of others, surprising phenomena can arise. For example, a high price of one asset, holding other prices constant, can cause agents to lower their expectation of that asset's payoff.

Grundy and McNichols (1989) and Brown and Jennings (1990) assume two trading periods and one risky asset. They show that uninformed traders learn about the asset payoff not only from current prices

[2] This conclusion does not extend to settings with imperfect competition, as we point out in Section 7.

but also from past ones because prices are noisy signals of asset payoffs. The optimal strategy thus uses the entire price history, in a manner similar to strategies used by technical traders. Wang (1993) studies a continuous-time setting with one risky asset. He shows that uninformed agents behave as price chasers, buying following a price increase. He also shows that return volatility and price reversal can be highest under asymmetric information than under full or no information. The latter result is consistent with our Proposition 6.5. Wang (1994) employs a similar model to study the behavior of trading volume and its relationship with price changes. These papers assume an asymmetric information structure. He and Wang (1995) study the joint behavior of trading volume and prices under a differential information structure. Vives (1995) studies the speed at which prices aggregate information under a combined asymmetric-differential information structure, where some agents observe conditionally independent signals about the payoff of a risky asset and the remaining agents observe no signal.

Much of the literature on REE with asymmetric information focuses on the informativeness of prices, rather than on market liquidity. Market liquidity is instead emphasized in a subsequent literature which combines asymmetric information with strategic behavior or sequential arrival of traders. This literature was pioneered by Glosten and Milgrom (1985) and Kyle (1985), and is surveyed in the next section. Yet, even REE models with asymmetric information have implications for market liquidity. We derive such implications in the context of our model in Propositions 6.4 and 6.5. Additional implications are derived, for example, in Eisfeldt (2004) and Cespa and Foucault (2011). Eisfeldt (2004) assumes that risk-averse entrepreneurs can sell stakes in projects on which they have private information. During times of high productivity, entrepreneurs undertake larger projects, and hence have a stronger motive to share risk. Thus, when productivity is high, adverse selection is low and liquidity is high. Cespa and Foucault (2011) show that asymmetric information can generate liquidity spillovers: because asset payoffs are correlated, a drop in liquidity in one asset reduces the information available on other assets, hence reducing the liquidity of those assets.

A number of recent papers examine whether agents require higher expected returns to invest in the presence of asymmetric information. O'Hara (2003) and Easley and O'Hara (2004) show that prices are lower and expected returns are higher when agents receive private signals than when signals are public. This comparison concerns Period 1 of our model, and reverses when using the alternative symmetric-information benchmark where no signals are observed. By contrast, we show (Proposition 6.6 and Vayanos and Wang (2012a)) that the price in Period 0 is lower under asymmetric information than under either symmetric-information benchmark. Comparing prices in Period 0 measures the ex-ante effect of the imperfection, i.e., what compensation do agents require to invest ex-ante knowing that they will face asymmetric information ex-post? Qiu and Wang (2010) derive similar results in an infinite-horizon model. Garleanu and Pedersen (2004) show in a model with risk-neutral agents and unit demands that asymmetric information can raise or lower prices, with the effect being zero when probability distributions are symmetric — as is the case under normality. Ellul and Pagano (2006) show that asymmetric information lowers prices in a model of IPO trading.

7

Imperfect Competition

In the perfect-market benchmark, agents are competitive and have no effect on prices. In many markets, however, some agents are large relative to others, in the sense that they can influence prices either because of their size or because of their information advantage. We model imperfect competition by assuming that some agents can exert market power in Period 1. We mainly focus on the case where liquidity demanders behave as a single monopolist, and consider more briefly monopolistic behavior by liquidity suppliers. We emphasize the former case because it has received more attention in the literature. When liquidity suppliers behave monopolistically, imperfect competition obviously influences the supply of liquidity. More surprisingly, it can also influence liquidity supply when liquidity demanders behave monopolistically and suppliers do not.

We consider both the cases where liquidity demanders have no private information on asset payoffs, and so information is symmetric, and where they observe the private signal (6.1), and so information is asymmetric. The second case nests the first by setting the variance σ_ϵ^2 of the signal noise to infinity. Hence, we treat both cases simultaneously, and compare imperfect competition to the competitive equilibrium with asymmetric information studied in Section 6.

The trading mechanism in Period 1 is that liquidity suppliers submit a demand function and liquidity demanders submit a market order, i.e., a price-inelastic demand function. Restricting liquidity demanders to trade by market order is without loss of generality: they do not need to condition their demand on price because they know all information available in Period 1.

7.1 Equilibrium

We conjecture that the price in Period 1 has the same affine form (6.2) as in the competitive case, with possibly different constants (a, b, c). Given (6.2), the demand function of liquidity suppliers is (6.5a) as in the competitive case. Substituting (6.5a) into the market-clearing equation (3.5), and using (6.4a), yields the price in Period 1 as a function of the liquidity demanders' market order θ_1^d:

$$S_1(\theta_1^d) = \frac{\bar{D} - \frac{\beta_\xi}{b} a + \frac{\alpha\sigma^2[D|S_1]}{1-\pi}(\pi\theta_1^d - \bar{\theta})}{1 - \frac{\beta_\xi}{b}}. \tag{7.1}$$

Liquidity demanders choose θ_1^d to maximize the expected utility

$$-\mathrm{E}\exp\left\{-\alpha\left[W_1 + \theta_1^d\left(D - S_1(\theta_1^d)\right) + z(D - \bar{D})\right]\right\}. \tag{7.2}$$

The difference with the competitive case is that liquidity demanders behave as a single monopolist and take into account the impact of their order θ_1^d on the price S_1. Proposition 7.1 characterizes the solution to the liquidity demanders' optimization problem.

Proposition 7.1. The liquidity demanders' market order in Period 1 satisfies

$$\theta_1^d = \frac{\mathrm{E}[D|s] - S_1(\theta_1^d) - \alpha\sigma^2[D|s]z + \hat{\lambda}\bar{\theta}}{\alpha\sigma^2[D|s] + \hat{\lambda}}, \tag{7.3}$$

where $\hat{\lambda} \equiv \frac{dS_1(\theta_1^d)}{d\theta_1^d} = \frac{\alpha\pi\sigma^2[D|S_1]}{(1-\pi)\left(1-\frac{\beta_\xi}{b}\right)}$.

Equation (7.3) determines θ_1^d implicitly because it includes θ_1^d in both the left- and the right-hand side. We write θ_1^d in the form (7.3) to facilitate the comparison with the competitive case. Indeed, the competitive counterpart of (7.3) is (6.5b), and can be derived by setting $\hat{\lambda}$ to zero. The parameter $\hat{\lambda}$ measures the price impact of liquidity demanders, and is closely related to the price impact λ. Because in equilibrium $\hat{\lambda} > 0$, the denominator of (7.3) is larger than that of (6.5b), and therefore θ_1^d is less sensitive to changes in $\mathrm{E}[D|s] - S_1$ and z than in the competitive case. Intuitively, because liquidity demanders take price impact into account, they trade less aggressively in response to their signal and their liquidity shock.

Substituting (6.5a) and (7.3) into the market-clearing equation (3.5), and proceeding as in Section 6, we find a system of three equations in (a, b, c). Proposition 7.2 solves this system.

Proposition 7.2. The price in Period 1 is given by (6.2), where

$$b = \frac{\pi\beta_s\sigma^2[D|S_1] + (1-\pi)\beta_\xi\sigma^2[D|s]}{2\pi\sigma^2[D|S_1] + (1-\pi)\sigma^2[D|s]}, \tag{7.4}$$

and (a, c) are given by (6.7a) and (6.7c), respectively. The linear equilibrium exists if $\sigma_\epsilon^2 > \hat{\sigma}_\epsilon^2$, where $\hat{\sigma}_\epsilon^2$ is the positive solution of

$$\alpha^2\hat{\sigma}_\epsilon^4\sigma_z^2 = \sigma^2 + \hat{\sigma}_\epsilon^2. \tag{7.5}$$

The price in the competitive market in Period 0 can be determined through similar steps as in Sections 3 and 6.

Proposition 7.3. The price in Period 0 is given by (3.10), where

$$M = \exp\left(\frac{1}{2}\alpha\Delta_2\bar{\theta}^2\right)\sqrt{\frac{1+\Delta_0\pi^2}{1+\Delta_0\left(1+\frac{2\hat{\lambda}}{\alpha\sigma^2[D|s]}\right)(1-\pi)^2 - \alpha^2\sigma^2\sigma_z^2}}, \tag{7.6}$$

$$\Delta_1 = \frac{\alpha^3 b\sigma^2(\sigma^2+\sigma_\epsilon^2)\sigma_z^2}{1+\Delta_0\left(1+\frac{2\hat{\lambda}}{\alpha\sigma^2[D|s]}\right)(1-\pi)^2-\alpha^2\sigma^2\sigma_z^2}, \tag{7.7a}$$

$$\Delta_2 = \frac{\alpha^3\sigma^4\sigma_z^2\left[1+\frac{\alpha(\beta_s-b)^2(\sigma^2+\sigma_\epsilon^2)\left(\alpha\sigma^2[D|s]+2\hat{\lambda}\right)}{\left(\alpha\sigma^2[D|s]+\hat{\lambda}\right)^2}\right]}{1+\Delta_0\left(1+\frac{2\hat{\lambda}}{\alpha\sigma^2[D|s]}\right)(1-\pi)^2-\alpha^2\sigma^2\sigma_z^2}, \tag{7.7b}$$

and Δ_0 is given by (6.8a).

7.2 Imperfect Competition and Illiquidity

We next examine how imperfect competition by liquidity demanders impacts the illiquidity measures and the expected return.

Proposition 7.4. Price impact λ is given by (6.9). It is the same as under competitive behavior when information is symmetric, and higher when information is asymmetric.

When information is asymmetric, imperfect competition lowers liquidity, as measured by price impact, even though liquidity suppliers are competitive. The intuition is that when liquidity demanders take into account their effect on price, they trade less aggressively in response to their signal and their liquidity shock. This reduces the size of both information- and liquidity-generated trades (hence lowering b in (6.9)). The relative size of the two types of trades (measured by c) remains the same, and so does price informativeness, measured by the signal-to-noise ratio. Monopoly trades thus have the same informational content as competitive trades, but are smaller in size. As a result, the signal per unit trade is higher, and so is the price impact λ of trades. Imperfect competition has no effect on price impact when information is symmetric because trades have no informational content.

An increase in information asymmetry, through a reduction in the variance σ_ϵ^2 of the signal noise, generates an illiquidity spiral. Because illiquidity increases, liquidity demanders scale back their trades. This raises the signal per trade size, further increasing illiquidity. When information asymmetry becomes severe, illiquidity becomes infinite and trade ceases, leading to a market breakdown. This occurs when $\sigma_\epsilon^2 \leq \hat{\sigma}_\epsilon^2$, i.e., for values of σ_ϵ^2 such that the equilibrium of Proposition 7.2 does not exist. Note that in our model non-competitive behavior is essential for the non-existence of an equilibrium with trade because such an equilibrium always exists under competitive behavior.

While imperfect competition raises price impact λ, it lowers price reversal γ. Intuitively, price reversal arises because the liquidity demanders' trades in Period 1 cause the price to deviate from fundamental value. Under imperfect competition, these trades are smaller and so is price reversal.

Proposition 7.5. Price reversal γ is given by (6.10), and is lower than under competitive behavior.

Imperfect competition can lower or raise the illiquidity discount. Indeed, since liquidity demanders scale back their trades, they render the price less responsive to their liquidity shock. Therefore, they can obtain better insurance against the shock, and become less averse to holding the asset in Period 0. This effect drives the illiquidity discount, and hence the asset's expected return, below the competitive value when information is symmetric. When information is asymmetric, the comparison can reverse. This is because the scaling back of trades generates the spiral of increasing illiquidity, and this reduces the insurance received by liquidity demanders.

Proposition 7.6. The asset's expected return $E(R)$ is lower than under perfect competition when information is symmetric, but can be higher when information is asymmetric.

The case where liquidity suppliers collude can be treated in a manner similar to the case where demanders collude, so we provide a brief sketch. Suppose that demanders are competitive but suppliers behave as a single monopolist in Period 1. Since suppliers do not know the liquidity shock z and signal s, their trading strategy is to submit a price-elastic demand function (rather than a market order). Non-competitive behavior renders this demand function less price-elastic than its competitive counterpart (6.5a). The lower elasticity manifests itself through an additive positive term in the denominator of the competitive demand (6.5a), exactly as is the case for liquidity demanders in (6.5b) and (7.3).

Because liquidity suppliers submit a less price-elastic demand function than in the competitive case, the trades of liquidity demanders have larger price impact. Hence, price impact λ and price reversal γ are larger than in the competitive case. The illiquidity discount is also larger because liquidity demanders receive worse insurance against the liquidity shock. Thus, imperfect competition by suppliers has the same effect as by demanders on λ, the opposite effect on γ, and the same or opposite effect on the illiquidity discount.

7.3 Literature

Two seminal papers on imperfect competition in financial markets and its relationship with asymmetric information are Kyle (1985, 1989). Kyle (1989) assumes a combined asymmetric-differential information structure, where some agents observe conditionally independent signals about the payoff of a risky asset and the remaining agents observe no signals. Agents submit demand functions, as in competitive rational expectations equilibrium (REE), but the equilibrium concept is instead Nash equilibrium in demand functions, as in Wilson (1979) and Klemperer and Meyer (1989). Noise traders add a stochastic amount to the asset supply, preventing prices from being fully revealing. Because informed agents take into account their effect on price, they trade less aggressively in response to their signal. Imperfect competition thus reduces the size of information-based trades. Since it has no effect on liquidity-generated trades, which are initiated by the exogenous noise traders, it lowers price informativeness.

Kyle (1985) assumes a risk neutral insider who observes a private signal about the payoff of a risky asset and can trade with market makers and noise traders. The insider and the noise traders submit market orders, which are batched together and absorbed by market makers. Because the latter are risk neutral and competitive, they compete a la Bertrand and absorb all orders at a price equal to their conditional expectation of the asset payoff. Imperfect competition reduces price informativeness, as in Kyle (1989). An advantage of Kyle (1985) is that it is highly tractable and can be extended in many directions. One important extension, performed in Kyle (1985), is to allow trading to take place dynamically, over more than one period. The insider then reveals his information slowly over time, as revealing it quickly would subject him to a higher price impact in the early periods. In the continuous-time limit, the insider reveals his information in a way that exactly equates price impact across time.

The model of Kyle (1985) has been extended in many other directions as well. A first extension is to introduce multiple insiders. Admati and Pfleiderer (1988) show that liquidity traders can concentrate their trades in the same period, to reduce price impact, and this effect can be amplified when there are multiple insiders. Holden and Subrahmanyam (1992) assume multiple insiders who receive a common signal about the payoff of a risky asset, and show that they reveal it almost immediately in the continuous-trading limit because each insider tries to exploit his information before others do. Foster and Viswanathan (1996) assume multiple insiders who receive imperfectly correlated signals, and show that information revelation slows down because of a "waiting-game" effect, whereby each insider attempts to learn the others' signals. Back et al. (2000) formulate this problem in continuous time and show that information is not fully revealed in prices until the end of the trading session.

A second extension is to drop the noise traders and derive non-informational trading from utility maximization. Glosten (1989) generates non-informational trading through a random endowment received by the insider. We make the same assumption, and the equilibrium in our Period 1 is closely related to the one that Glosten (1989) derives in the case where market makers are competitive. Glosten (1989) assumes risk-neutral market makers; a paper even closer to our model is Bhattacharya and Spiegel (1991), which assumes that liquidity suppliers are risk averse.[1] Both papers find that the market breaks

[1] Both papers assume one trading period and do not derive effects on ex-ante expected returns, as we do.

down when information asymmetry is severe. The mechanism is the same as in our model, and the key assumptions are that some liquidity demanders are informed and all are non-price-takers.[2] The idea that adverse selection can cause market breakdown dates back to Akerlof (1970).

Other extensions are to introduce non-normal probability distributions for the asset payoff, combine risk aversion with dynamics, and allow for a minimum trade size. Back (1992) shows that the result on the equalization of price impact across time extends to general payoff distributions. Holden and Subrahmanyam (1994) and Baruch (2002) show that a risk averse insider reveals his information faster than a risk neutral one because he is eager to reduce the uncertainty at which his trades will be executed. Back and Baruch (2004) assume that noise traders execute discrete transactions rather than trading continuously, in which case the insider must do the same so not to be revealed. They show that the insider follows a mixed strategy, and can trade in a direction opposite to his information in some cases.

A further extension is to change the information structure. Kyle (1985) assumes that the insider receives all his information in an initial period, and the information is announced publicly in a final period. Chau and Vayanos (2008) and Caldentey and Stacchetti (2010) show that when the insider receives a constant flow of new information over time, he chooses to reveal it infinitely fast in the continuous-trading limit. This result is in sharp contrast to Kyle (1985), where revelation is slow. Moreover, markets are arbitrarily close to efficiency and yet informed traders earn abnormal profits, in sharp contrast to Grossman and Stiglitz (1980). Efficient markets and insider profits are not contradictory because continuous trading gives insiders the flexibility to earn profits even though they reveal each piece of new information within a very short interval.[3] Other models exploring insider trading with a

---

[2] For example, market breakdown does not occur in Kyle (1985) because noise traders submit price-inelastic demands, which can be viewed as an extreme form of price taking.

[3] Jackson (1991) provides an alternative resolution of the Grossman and Stiglitz (1980) paradox within a static setting. He assumes that agents can acquire private signals at a cost and submit demand functions as in Kyle (1989). Unlike in Kyle (1989), there are no noise traders. The equilibrium price is fully revealing and yet agents have an incentive to acquire information. This is because information helps them predict their price impact,

flow of new information are Back and Pedersen (1998) and Guo and Ou-Yang (2010).

A final set of extensions examine issues relating to market design. For example, Chowhdry and Nanda (1991) study the competition between market venues. Fishman and Hagerty (1992), Leland (1992) and Repullo (1999) study whether insider trading is desirable or should be banned. Admati and Pfleiderer (1991) study "sunshine trading," whereby liquidity traders pre-announce their intention to trade, so to distinguish themselves from insiders and reduce their trading costs. Pagano and Roell (1996) and Naik et al. (1999) study the effects of a transparency regulation requiring disclosure of all trades but of not traders' identities. Huddart et al. (2001) consider instead a regulation requiring disclosure of trades by insiders. They show that the regulation speeds up information revelation and reduces insiders' profits. It also induces the insiders to trade less aggressively and follow a mixed strategy, trading occasionally in a direction opposite to their information. Buffa (2011) shows that because of the latter effect, the regulation can instead slow down information revelation when insiders are risk averse.

Kyle (1985) and much of the subsequent literature assume that the non-price-taking agents are insiders who receive private information about asset payoffs. In many cases, however, agents without such information affect prices simply because of the size of their trades. For example, trades by pension funds can exceed the average daily volume of many stocks, and are often triggered by reasons other than information, e.g., regulatory constraints. Vayanos (1999) assumes that large traders with no private information about asset payoffs receive random endowments over time and need to share risk. He shows that these agents break their trades into small pieces so to reduce price impact, and risk sharing is accomplished slowly even in the continuous-trading limit. What deters them from trading faster is that this will signal to the market that a larger trade is yet to come, and so price impact will

---

which the price does not reveal. Normal-linear models cannot generate this effect because price impact is a constant independent of information. For an analysis of information revelation without noise traders, normality and linearity, see also Laffont and Maskin (1990).

be large. Vayanos (2001) shows that the presence of noise traders in this setting can accelerate trading and hence improve risk sharing.[4]

Large traders who trade over time to share risk are similar to durable-good monopolists. According to the Coase conjecture, the monopolists should trade infinitely fast in the continuous-trading limit. Trading occurs slowly in Vayanos (1999) because each trader is the only one to observe his endowment and hence his eagerness to share risk; if instead endowments are publicly observed, the Coase conjecture holds. DeMarzo and Urosevic (2006) consider a general setting where a large trader not only needs to share risk but can also take actions to increase asset payoffs, e.g., monitor the firm's managers. The trader's eagerness to share risk is public information. DeMarzo and Urosevic (2006) confirm the Coase conjecture in the case where asset payoffs are independent of the trader's actions. Rostek and Weretka (2011) study risk sharing in a dynamic setting where agents' endowments are public information. They decompose the price impact of a trade into a permanent component, due to the risk aversion of agents taking the other side, and a temporary one, due to their monopoly power.

When large traders affect prices, information about their future trades is valuable to others. This is so even when large traders themselves have no information about asset payoffs. Cao et al. (2006) label information about future large trades "inventory information." Brunnermeier and Pedersen (2005) assume that a large trader needs to sell because of financial distress, and show that other traders exploit this information by selling at the same time as him. Such "predatory" behavior benefits these traders because they cause the distressed trader to sell at low prices, at which they can buy. Pritsker (2005) studies predatory behavior in a multi-asset setting. Attari et al. (2005), Fardeau (2011) and Venter (2011) model the financial constraints of distressed traders and examine how predatory behavior by others can bring them closer to distress by moving asset prices against them.

[4] Bertsimas and Lo (1998), Almgren and Chriss (1999), Almgren (2003) and Huberman and Stanzl (2005) study the optimal policy of large traders in partial-equilibrium settings, under exogenous price dynamics. Obizhaeva and Wang (2006) derive the price dynamics faced by large traders from a model of the limit-order book, which describes how new limit orders arrive after existing ones are consumed.

Carlin et al. (2007) derive predatory behavior as a breakdown of collaboration in a repeated game.

Most papers mentioned so far emphasize non-price-taking behavior by liquidity demanders; liquidity suppliers, such as market makers, are assumed to behave competitively. Biais (1993) studies how oligopolistic market makers bid for an order, depending on whether or not they know the inventories of their competitors. He relates the quality of market makers' information to whether the market is centralized or fragmented. Earlier papers on oligopolistic market makers include Ho and Stoll (1980) and Copeland and Galai (1983).

Glosten (1989) shows that when information is asymmetric, a market with a monopoly market maker can dominate one with competitive market makers. This is because the market can break down with competitive market makers, but breakdown can be avoided with a monopoly market maker. Glosten (1989) models perfect competition between market makers in terms of a zero-profit condition. Glosten (1994) derives this condition as the equilibrium of a game in which market makers post price–quantity schedules. Bernhardt and Hughson (1997) study this game in the case of two oligopolistic market makers. Biais et al. (2000) study the game for a general number of market makers and provide a full characterization of the equilibrium. For a finite number of market makers the equilibrium has a Cournot flavor, and it converges to the competitive case characterized by Glosten (1989) as the number goes to infinity. Back and Baruch (2011) provide an alternative characterization of the same game. Liu and Wang (2012) assume that market makers are risk averse and compete by posting quantities rather than price–quantity schedules.

Models with non-price-taking behavior study the interaction between small numbers of agents. In this sense, they are related to models of sequential order arrival, in which traders arrive in the market one at a time and remain there for a short period. The latter models assume implicitly participation costs since agents are not present in the market until when they arrive. Early models in that spirit include Garman (1976), Amihud and Mendelson (1980) and Ho and Stoll (1981), in which market makers receive an exogenous flow of orders.

Glosten and Milgrom (1985) propose a highly tractable model of sequential order arrival with asymmetric information. Some of the agents receive a private signal about the asset payoff, while others do not and trade for liquidity reasons. Upon arriving in the market, agents can execute a buy or sell transaction of a fixed size with market makers. As in Kyle (1989), market makers are risk neutral and competitive. Therefore, they compete a la Bertrand and absorb orders at a price equal to their conditional expectation of the asset payoff. In equilibrium, the bid price that market makers quote to buy from other agents is lower than the ask price that they quote to sell to them. This is because when market makers buy, they suspect that other agents might have sold to them because of negative information. Glosten and Milgrom (1985) thus link the bid-ask spread to asymmetric information, building on earlier works by Bagehot (1971) and Copeland and Galai (1983).

The models of Glosten and Milgrom (1985) and Kyle (1985) give rise to different measures of illiquidity. Illiquidity in Glosten and Milgrom (1985) is measured by the bid-ask spread since all transactions are assumed to be of a fixed size. By contrast, in Kyle (1985) transactions can be of any size since probability distributions are normal and trading strategies are linear. Illiquidity is measured by the sensitivity of price to quantity, which corresponds to λ in our model. While the bid-ask spread and λ are derived within different models, the share the basic property of being increasing in the degree of asymmetric information. Easley and O'Hara (1987) consider a hybrid model in which agents arrive in the market one at a time and can execute transactions of variable size with market makers. The prices that market makers post depend on quantity, in a spirit similar to Kyle (1985). Easley and O'Hara (1992) allow the time when private information arrives to be stochastic. They show that the bid-ask spread increases following a surge in trading activity because market makers infer that private information has arrived.

A recent literature studies sequential order arrival in limit-order markets, where there are no designated market makers and liquidity is supplied by the arriving agents. Agents can execute a buy or sell transaction of a fixed size. Impatient agents execute this transaction immediately upon arrival through a market order, and hence demand liquidity. Patient agents submit instead a limit order, i.e., a price-elastic demand function, which is executed against future market orders. Hence, they supply liquidity to future agents. Papers in that literature include Parlour (1998), Foucault (1999), Foucault et al. (2005), Goettler et al. (2005) and Rosu (2009).[5] These papers determine the bid-ask spread that results from the submitted limit orders, the choice of agents between market and limit orders, the expected time for limit orders to execute, etc. A positive bid-ask spread arises even in the absence of asymmetric information, and is decreasing in the degree of competition between limit-order suppliers. This parallels our result that λ is larger when liquidity suppliers behave monopolistically than when they are competitive.

[5] These papers assume that agents have market power. Biais et al. (2011) assume instead that agents are competitive and observe their valuation for an asset only infrequently. They show that the optimal orders that agents submit at the observation times can be price-contingent and concern future execution.

8

Funding Constraints

Agents' portfolios often involve leverage, i.e., borrow cash to establish a long position in a risky asset, or borrow a risky asset to sell it short. In the perfect-market benchmark, agents can borrow freely provided that they have enough resources to repay the loan. But as the Corporate Finance literature emphasizes, various frictions can limit agents' ability to borrow and fund their positions. These frictions can also influence the supply of liquidity in the market.

Since in our model consumption is allowed to be negative and unbounded from below, agents can repay a loan of any size by reducing consumption. Negative consumption can be interpreted as a costly activity that agents undertake in Period 2 to repay a loan. We derive a funding constraint by assuming that agents cannot commit to reduce their consumption below a level $-A \leq 0$. This nests the case of full commitment assumed in the rest of this paper ($A = \infty$), and the case where agents can walk away from a loan rather than engaging in negative consumption ($A = 0$). Because our focus is on how the funding constraint influences the supply of liquidity, we impose it on liquidity suppliers only, i.e., assume that the lack of commitment concerns only them.

For simplicity, we assume that loans must be fully collateralized in the sense that agents must be able to commit enough resources to cover any losses in full. To ensure that full collateralization is possible, we replace normal distributions by distributions with bounded support. We denote the support of the asset payoff D by $[\bar{D} - b_D, \bar{D} + b_D]$ and that of the liquidity shock z by $[-b_z, b_z]$. We assume that D and z are distributed symmetrically around their respective means, D is positive (i.e., $\bar{D} - b_D \geq 0$), and agents receive a positive endowment B of the riskless asset in Period 0.

8.1 Equilibrium

In Period 1, a liquidity demander chooses holdings θ_1^d of the risky asset to maximize the expected utility (3.1). The expectation over D is

$$-\exp\left\{-\alpha\left[W_1 + \theta_1^d(\bar{D} - S_1) - f(\theta_1^d + z)\right]\right\}, \tag{8.1}$$

where

$$f(\theta) \equiv \frac{\log \mathrm{E} \exp\left[-\alpha\theta(D - \bar{D})\right]}{\alpha}. \tag{8.2}$$

Equation (8.1) generalizes (3.2), derived under normality, to any symmetric distribution. The function $f(\theta)$, equal to $\frac{1}{2}\alpha\theta^2$ under normality, is positive, symmetric around the y-axis, and convex.[1] Maximizing (8.1) over θ_1^d yields the demand function

$$\theta_1^d = \left(f'\right)^{-1}(\bar{D} - S_1) - z. \tag{8.3}$$

Since $f(\theta)$ is convex, the demand θ_1^d is a decreasing function of the price S_1.

A liquidity supplier chooses holdings θ_1^s of the risky asset to maximize the expected utility

$$-\exp\left\{-\alpha\left[W_1 + \theta_1^s(\bar{D} - S_1) - f(\theta_1^s)\right]\right\}, \tag{8.4}$$

which can be derived from (8.1) by setting $z = 0$. The optimization is subject to a funding constraint. Indeed, losses from investing in the

[1] The function $\alpha f(\theta)$ is the cumulant-generating function of $-\alpha(D - \bar{D})$. Cumulant-generating functions are convex. Symmetry follows because D is distributed symmetrically around $\bar{D}$. Positivity follows from $f(0) = 0$, symmetry and convexity.

risky asset can be covered by wealth W_1 or negative consumption. Since suppliers must be able to cover losses in full, and cannot commit to consume less than $-A$, losses cannot exceed $W_1 + A$, i.e.,

$$\theta_1^s(S_1 - D) \leq W_1 + A, \quad \text{for all } D.$$

This yields the constraint

$$m|\theta_1^s| \leq W_1 + A, \tag{8.5}$$

where

$$m \equiv S_1 - \min_D D, \quad \text{if } \theta_1^s > 0, \tag{8.6a}$$

$$m \equiv \max_D D - S_1, \quad \text{if } \theta_1^s < 0. \tag{8.6b}$$

The constraint (8.5) requires that a position of θ_1^s shares is backed by capital $m|\theta_1^s|$. This limits the size of the position as a function of the capital $W_1 + A$ available to suppliers in Period 1. Suppliers' capital is the sum of the capital W_1 that they physically own in Period 1, and the capital A that they can access through their commitment to consume — A in Period 2. The parameter m is the required capital per share of levered position, and can be interpreted as a margin or haircut. The margin is equal to the maximum possible loss per share. For example, the margin (8.6a) for a long position does not exceed the asset price S_1, and is strictly smaller if the asset payoff D has a positive lower bound (i.e., $\min_D D = \bar{D} - b_D > 0$).[2]

Intuitively, the constraint (8.5) can bind when there is a large discrepancy between the price S_1 and the expected payoff $\bar{D}$, since this is when liquidity suppliers want to hold large positions. There is, however, a countervailing effect because of a decrease in the margin. When, for example, S_1 is low, suppliers want to hold large long positions, but the margin is small because the maximum possible loss is small. The required capital (position size times margin) increases in the discrepancy between S_1 and $\bar{D}$ under the sufficient condition

$$2\alpha\pi b_D b_z < 1, \tag{8.7}$$

which for simplicity we assume from now on.

[2] The margins (8.6a) for a long position and (8.6b) for a short position are finite because D has bounded support. Our analysis can accommodate short-sale constraints, i.e., infinite margins for short positions, by setting the upper bound of D to infinity.

Proposition 8.1. The equilibrium in Period 1 has the following properties:

- The funding constraint (8.5) never binds if
$$B + A + \bar{\theta}(\bar{D} - b_D) - \pi b_z \left[b_D - f'(\bar{\theta} + \pi b_z)\right] \geq 0. \tag{8.8}$$
Otherwise, (8.5) binds for $z \in [-b_z, -\overline{z}) \cup (\underline{z}, b_z]$, where $0 < \underline{z} < \overline{z} \leq b_z$.
- An increase in z lowers the price S_1 and raises the liquidity suppliers' position θ_1^s. When (8.5) does not bind, $\theta_1^s = \bar{\theta} + \pi z$ and
$$S_1 = \bar{D} - f'(\bar{\theta} + \pi z). \tag{8.9}$$

The funding constraint never binds if agents receive a large endowment B of the riskless asset in Period 0, or if they can commit to a large negative consumption $-A$ in Period 2. In both cases, the capital that they can access in Period 1 is large. If instead B and A are small, the constraint binds for large positive and possibly large negative values of the liquidity shock z. For example, when z is large and positive, the price S_1 is low and liquidity suppliers are constrained because they want to hold large long positions. Setting

$$K^* \equiv \pi b_z \left[b_D - f'(\bar{\theta} + \pi b_z)\right] - \bar{\theta}(\bar{D} - b_D),$$

we refer to the region $B + A > K^*$, where liquidity suppliers are well-capitalized and the constraint never binds, as the *abundant-capital* region, and to the region $B + A < K^*$, where the constraint binds for some values of z, as the *scarce-capital* region. Note that in both regions, the constraint does not bind for $z = 0$. Indeed, the unconstrained outcome for $z = 0$ is that liquidity suppliers maintain their endowments $\bar{\theta}$ of the risky asset and B of the riskless asset. Since this yields positive consumption, the constraint is met.

An increase in the liquidity shock z lowers the price S_1 and raises the liquidity suppliers' position θ_1^s. These results are the same as in the perfect-market benchmark of Section 3, but the intuition is more complicated when the funding constraint binds. Suppose that capital is scarce (i.e., $B + A < K^*$), and z is large and positive, in which case suppliers hold long positions and are constrained. The intuition why they can buy more, despite the constraint, when z increases is as follows. Since the price S_1 decreases, suppliers realize a capital loss on the $\bar{\theta}$ shares of the risky asset that they carry from Period 0. This reduces their wealth in Period 1 and tightens the constraint. At the same time, a decrease in S_1 triggers an equal decrease in the margin (8.6a) for long positions, and loosens the constraint. This effect is equivalent to a capital gain on the θ_1^s shares that suppliers hold in Period 1. Because suppliers are net buyers for $z > 0$ (i.e., $\theta_1^s > \bar{\theta}$), the latter effect dominates, and suppliers can buy more in response to an increase in z.

To determine the price in Period 0, we make the simplifying assumption that the risk-aversion coefficient α is small. We denote by (σ^2, σ_z^2) the variances of (D, z), by $k \equiv [\mathrm{E}[D - \bar{D}]^4/\sigma^4] - 3$ the curtosis of D, by $F(z)$ the cumulative distribution function of z, and by $o(\alpha^n)$ terms smaller than α^n.

Proposition 8.2. Suppose that α is small. The price in Period 0 is

$$S_0 = \bar{D} - \alpha\sigma^2\bar{\theta} - \alpha^3\sigma^4\left[\left(1 + \frac{1}{2}k\right)\sigma_z^2\pi^2 + \frac{1}{6}k\bar{\theta}^2\right]\bar{\theta} + o(\alpha^3) \quad (8.10)$$

when capital is abundant, and

$$\begin{aligned} S_0 = {} & \bar{D} - \alpha\sigma^2\bar{\theta} - \alpha\sigma^2(1 - \pi) \\ & \times \left[\int_{\underline{z}}^{\overline{z}} (z - \underline{z})dF(z) + \int_{\overline{z}}^{b_z} (\overline{z} - \underline{z})dF(z)\right] + o(\alpha) \end{aligned} \quad (8.11)$$

when capital is scarce.

8.2 Funding Constraints and Illiquidity

We next examine how the funding constraint impacts the illiquidity measures and the expected return. We compute these variables in the abundant-capital region (liquidity suppliers are well-capitalized and unconstrained by leverage for all values of the liquidity shock z), and compare with the scarce-capital region.

Proposition 8.3. Suppose that α is small or z is drawn from a two-point distribution. Price impact λ is higher when capital is scarce than when it is abundant.

Proposition 8.4. Price reversal γ is higher when capital is scarce than when it is abundant.

The intuition is as follows. When the liquidity shock z is close to zero, the constraint does not bind in both the abundant- and scarce-capital regions, and therefore price and volume are identical in the two regions. For larger values of z, the constraint binds when capital is scarce, impairing suppliers' ability to accommodate an increase in z. As a result, an increase in z has a larger effect on price and a smaller effect on volume. Since the effect on price is larger, so is the price reversal γ. Price impact λ is also larger because it measures the price impact per unit of volume. Note that λ measures an average price impact, i.e., the average slope of the relationship between price change and signed volume. This relationship exhibits an important non-linearity when capital is scarce: the slope increases for large values of z, which is when the constraint binds. This property distinguishes funding constraints from other imperfections.

The illiquidity discount, and hence the asset's expected return, is higher when capital is scarce. This is because the funding constraint binds asymmetrically: it is more likely to bind when liquidity demanders sell ($z > 0$) than when they buy ($z < 0$). Indeed, the constraint binds when the suppliers' position is large in absolute value — and a large position is more likely when suppliers buy in Period 1 because this adds to the long position $\bar{\theta}$ that they carry from Period 0. Since price movements in Period 1 are exacerbated when the constraint binds, and the constraint is more likely to bind when demanders sell, the average price in Period 1 is lower when capital is scarce. This yields a lower price in Period 0.

Proposition 8.5. Suppose that α is small. The asset's expected return $E(R)$ is lower when capital is scarce than when it is abundant.

8.3 Literature

The literature on funding constraints in financial markets can be viewed as part of a broader literature on the limits of arbitrage. Indeed, both literatures emphasize the idea that some traders rely on external capital, which is costlier than internal capital, and this affects liquidity and asset prices. External capital can take the form of collateralized debt, as in our model, or other forms such as equity. We first survey work that derives funding constraints from collateralized debt, and then survey more briefly the broader theoretical literature on the limits of arbitrage. An extensive survey of the latter literature is Gromb and Vayanos (2010).

The effects of funding constraints have been studied in macroeconomic settings, starting with Bernanke and Gertler (1989) and Kiyotaki and Moore (1997). In these papers, adverse shocks to economic activity depress the collateral values of productive assets, and this reduces lending and amplifies the drop in activity. Similar amplification effects arise in financial-market settings, as we point out below.

A number of papers link the tightness of funding constraints to the volatility of the collateral. Hart and Moore (1994, 1995) show that uncertainty about assets' liquidation values impairs agents' ability to borrow. Shleifer and Vishny (1992) endogenize liquidation values and the ability to borrow in market equilibrium. Geanakoplos (1997, 2003) defines collateral equilibrium, in which agents borrow to buy financial assets and post the assets as collateral. The amount of collateral is determined endogenously in equilibrium, and is increasing in asset volatility. Moreover, if volatility increases following adverse shocks, funding constraints tighten, and this causes agents to sell assets, amplifying the shocks. The link between volatility and ability to borrow is also present in our model. Indeed, an increase in the parameter b_D, which measures the dispersion of the asset payoff distribution, raises the margins in (8.6a) and (8.6b), holding the price S_1 constant. The funding constraint (8.5) in our model is derived along the lines of Geanakoplos (2003), who also provides conditions under which full collateralization is an equilibrium outcome.

Gromb and Vayanos (2002) link market liquidity to the capital of financial intermediaries and their funding constraints — a link that we also derive in Propositions 8.3 and 8.4. Investors are subject to liquidity shocks and can realize gains from trade across segmented markets by trading with intermediaries. Intermediaries exploit price discrepancies, and in doing so supply liquidity to investors: they buy low in a market where investors are eager to sell, and sell high in a market where investors are eager to buy, thus supplying liquidity to both sets of investors. Intermediaries fund their position in each market using collateralized debt, and face a funding constraint along the lines of (8.5). Shocks to asset prices that trigger capital losses by intermediaries, tighten the intermediaries' funding constraints and force them to reduce their positions. This lowers market liquidity and amplifies the shocks.

Amplification effects do not arise in our model because liquidity suppliers increase their position θ_1^s in Period 1 following an increase in the liquidity shock z. Amplification effects require instead that suppliers decrease their position, hence becoming demanders of liquidity. Recall that suppliers in our model are able to increase their position following an increase in z because while their wealth decreases, there is a stronger countervailing effect caused by a decrease in the margin. Amplification effects arise when the margin instead increases, as in Geanakoplos (1997, 2003), or stays constant. They can arise even when the margin decreases but there are multiple periods, as in Gromb and Vayanos (2002). Kondor (2009) shows that amplification effects can arise even in the absence of shocks. Indeed, if a price discrepancy between two assets were to remain constant or decrease over time, intermediaries would exploit it and reduce it to a level from which it could increase.

In Gromb and Vayanos (2002) and Kondor (2009), intermediaries have one investment opportunity, which is a long–short position involving assets trading in segmented markets. Some papers study the effects of funding constraints when intermediaries have multiple investment opportunities. Brunnermeier and Pedersen (2009) show in a static setting that funding constraints generate not only amplification but also contagion, whereby shocks to one investment opportunity are transmitted to otherwise unrelated opportunities through changes in intermediaries' positions. Moreover, a tightening of funding constraints has the largest impact on the prices of more volatile opportunities because these require more collateral. Gromb and Vayanos (2011a,b) derive the joint dynamics of intermediary capital, asset volatility, correlations and liquidity. They show that amplification and contagion are stronger when intermediary capital is neither too high nor too low. Other papers that derive amplification and contagion from funding constraints include Aiyagari and Gertler (1999), Allen and Gale (2000), Geanakoplos and Fostel (2008), Pavlova and Rigobon (2008), Adrian et al. (2009), Rytchkov (2011), Danielsson et al. (2012), and Chabakauri (2012).[3]

Funding constraints can give rise to price discrepancies between assets with identical payoffs, even in the absence of market segmentation. Basak and Croitoru (2000) show that a derivative can carry a price premium relative to a perfectly correlated underlying asset when agents are constrained in the size of their long position in the asset and cannot short-sell the derivative. This is because optimistic agents have a high demand for the derivative, in which their long positions are unconstrained, and pessimistic agents cannot accommodate this demand by shorting the derivative. Garleanu and Pedersen (2011) assume that the constraint on long positions takes the form of a margin requirement, which is smaller for the derivative. They show that the derivative's lower margin requirement causes it to carry a price premium relative to the underlying asset. Other papers examining violations of the law of one price generated by funding constraints include Cuoco (1997), Detemple and Murthy (1997), Geanakoplos (2003), and Basak and Croitoru (2006).

Liu and Longstaff (2004) study how funding-constrained intermediaries exploit price discrepancies under an exogenous price process. They show that a funding constraint, along the lines of (8.5), prevents

[3] Amplification and contagion can also be derived in models without explicit funding constraints but where risk aversion depends on wealth. This is done in Kyle and Xiong (2001) and Xiong (2001), who endow some agents with logarithmic utility, under which the coefficient of absolute risk aversion decreases in wealth. Following adverse shocks, these agents reduce their positions because they become more risk averse and not because they hit funding constraints. The analysis has similarities to that with funding constraints, e.g., amplification and contagion are stronger when the capital of the agents with logarithmic utility is neither too high nor too low. An important difference is in the welfare and policy implications: funding constraints can create inefficiencies and the scope for welfare-improving policies, while wealth effects preserve the Pareto optimality of equilibrium.

drastically intermediaries from exploiting opportunities that appear to be perfect arbitrages. Other papers on optimal portfolio policy under funding constraints are Grossman and Vila (1992), Jurek and Yang (2007) and Milbradt (2012).

Early work on the limits of arbitrage does not consider funding constraints explicitly, but argues that such constraints can shorten traders' horizons, and this can affect asset prices. De Long et al. (1990) show that short horizons can cause deviations from the law of one price. They assume an infinite-horizon economy, two assets with identical payoffs, and stochastic shocks to the demand for one of the assets. They show that when traders have short horizons there exist two equilibria: one in which the assets trade at the same price and one in which they trade at different prices. The intuition for the latter equilibrium is that agents do not trade aggressively against price discrepancies between the two assets for fear that they might widen in the short run. As a consequence, demand shocks can cause price discrepancies and render traders' belief self-fulfilling.

Tuckman and Vila (1992, 1993) show that short horizons can arise endogenously because of holding costs. Moreover, holding costs can render traders unwilling to exploit price discrepancies between assets with similar payoffs for fear that they might widen in the short run. Dow and Gorton (1994) assume short horizons and show that holding costs can generate large mispricings. Casamatta and Pouget (2011) endogenize short horizons based on moral hazard between fund managers and investors, and show that they cause prices to be less informative.

Shleifer and Vishny (1997) model the reliance of traders on external capital and its implications for traders' horizons and asset pricing. They assume that traders can buy a underpriced asset but run the risk that the mispricing might worsen in the short run. Traders can raise external funds to buy the asset, but the suppliers of the funds can request them back if the trade performs poorly in the short run. This assumed performance–flow relationship can generate amplification effects: following demand shocks that cause the mispricing to worsen in the short run, traders are deprived of funds and must sell the asset, causing the mispricing to worsen further.

Shleifer and Vishny (1997) derive the funding constraint from equity finance: traders can be interpreted as managers of an open-end fund raising equity from fund investors. Yet, the amplification effects that they find are similar to those in the papers that derive funding constraints from collateralized debt. Recent work on the limits of arbitrage seeks to derive funding constraints from optimal contracts, instead of assuming an exogenous contract form. Examples are Acharya and Viswanathan (2011), Hombert and Thesmar (2011), Biais et al. (2012), and He and Krishnamurthy (2012). Endogenizing the constraints would help identify whether the common results, such as amplification, are driven by a single underlying friction, or whether the constraints are fundamentally different. Recent work on the limits of arbitrage also seeks to develop tractable dynamic multi-asset models that can address empirical puzzles. The survey by Gromb and Vayanos (2010) provides more details and references.

Funding constraints can interact with other market imperfections. Yuan (2005) and Albagli (2011) consider the interaction with asymmetric information, and impose funding constraints on informed agents. Yuan (2005) shows that when prices drop, informed agents become constrained and hence prices become less informative. The resulting increase in uncertainty exacerbates the price drop, causing volatility to be asymmetric and higher on the downside. Albagli (2011) derives multiple equilibria, through a mechanism that is reminiscent of De Long et al. (1990) but does not require an infinite horizon. When future demand shocks are expected to have a large effect on prices, funding-constrained agents do not trade aggressively on their information. This makes prices less informative, hence reducing the willingness of future agents to absorb demand shocks. Cespa and Vives (2012) derive a similar mechanism in a setting where traders have short horizons.

9

Search

In the perfect-market benchmark, the market is organized as a centralized exchange. Many markets, however, have a more decentralized form of organization. For example, in over-the-counter markets, investors negotiate prices bilaterally with dealers. Locating suitable counter-parties in these markets can take time and involve search.

To model decentralized markets, we assume that agents do not meet in a centralized exchange in Period 1, but instead must search for counterparties. When a liquidity demander meets a supplier, they bargain bilaterally over the terms of trade, i.e., the number of shares traded and the share price. We assume that bargaining leads to an efficient outcome, and denote by $\phi \in [0,1]$ the fraction of transaction surplus appropriated by suppliers. We denote by N the measure of bilateral meetings between demanders and suppliers. This parameter characterizes the efficiency of the search process, and is bounded by $\min\{\pi, 1-\pi\}$ since there cannot be more meetings than demanders or suppliers. Assuming that all meetings are equally likely, the probability of a demander meeting a supplier is $\pi^d \equiv N/\pi$, and of a supplier meeting a demander is $\pi^s \equiv N/(1-\pi)$.

9.1 Equilibrium

Prices in Period 1 are determined through pairwise bargaining between liquidity demanders and suppliers. Agents' outside option is not to trade and retain their positions from Period 0, which in equilibrium are equal to $\bar{\theta}$. The consumption in Period 2 of a liquidity supplier who does not trade in Period 1 is $C_2^{sn} = W_0 + \bar{\theta}(D - S_0)$. This generates a certainty equivalent

$$\text{CEQ}^{sn} = W_0 + \bar{\theta}(\bar{D} - S_0) - \frac{1}{2}\alpha\sigma^2\bar{\theta}^2, \tag{9.1}$$

where the first two terms are the expected consumption, and the third a risk adjustment quadratic in position size. If the supplier buys x shares at price S_1, the certainty equivalent becomes

$$\text{CEQ}^{s} = W_0 + \bar{\theta}(\bar{D} - S_0) + x(\bar{D} - S_1) - \frac{1}{2}\alpha\sigma^2(\bar{\theta} + x)^2 \tag{9.2}$$

because the position becomes $\bar{\theta} + x$. Likewise, the certainty equivalent of a liquidity demander who does not trade in Period 1 is

$$\text{CEQ}^{dn} = W_0 + \bar{\theta}(\bar{D} - S_0) - \frac{1}{2}\alpha\sigma^2(\bar{\theta} + z)^2, \tag{9.3}$$

and if the demander sells x shares at price S_1, the certainty equivalent becomes

$$\text{CEQ}^{d} = W_0 + \bar{\theta}(\bar{D} - S_0) - x(\bar{D} - S_1) - \frac{1}{2}\alpha\sigma^2(\bar{\theta} + z - x)^2. \tag{9.4}$$

Under efficient bargaining, x maximizes the sum of certainty equivalents $\text{CEQ}^s + \text{CEQ}^d$. The maximization yields $x = z/2$, i.e., the liquidity shock is shared equally between the two agents. The price S_1 is such that the supplier receives a fraction ϕ of the transaction surplus, i.e.,

$$\text{CEQ}^{s} - \text{CEQ}^{sn} = \phi\left(\text{CEQ}^{s} + \text{CEQ}^{d} - \text{CEQ}^{sn} - \text{CEQ}^{dn}\right). \tag{9.5}$$

Proposition 9.1. When a supplier and a demander meet in Period 1, the supplier buys $z/2$ shares at the price

$$S_1 = \bar{D} - \alpha\sigma^2\left[\bar{\theta} + \frac{1}{4}z(1 + 2\phi)\right]. \tag{9.6}$$

Equation (9.6) implies that the impact of the liquidity shock z on the price in Period 1 increases in the liquidity suppliers' bargaining power ϕ. When, for example, $z > 0$, liquidity demanders need to sell, and greater bargaining power by suppliers results in a lower price. Comparing (9.6) to its centralized-market counterpart (3.6) reveals an important difference: price impact in the search market depends on the distribution of bargaining power within a meeting, characterized by the parameter ϕ, while price impact in the centralized market depends on aggregate demand-supply conditions, characterized by the measures $(\pi, 1-\pi)$ of demanders and suppliers. The price in the centralized market in Period 0 can be determined through similar steps as in previous sections.

Proposition 9.2. The price in Period 0 is

$$S_0 = \bar{D} - \alpha\sigma^2\bar{\theta} - \frac{\frac{N(1+\phi)}{2G_2^{\frac{3}{2}}}\exp\left(\frac{\alpha^4\sigma^4\sigma_z^2\bar{\theta}^2}{2G_2}\right) + \frac{\pi-N}{G_3^{\frac{3}{2}}}\exp\left(\frac{\alpha^4\sigma^4\sigma_z^2\bar{\theta}^2}{2G_3}\right)}{\frac{N}{\sqrt{G_1}} + 1 - \pi - N + \frac{N}{\sqrt{G_2}}\exp\left(\frac{\alpha^4\sigma^4\sigma_z^2\bar{\theta}^2}{2G_2}\right) + \frac{\pi-N}{\sqrt{G_3}}\exp\left(\frac{\alpha^4\sigma^4\sigma_z^2\bar{\theta}^2}{2G_3}\right)}\alpha^3\sigma^4\sigma_z^2\bar{\theta}, \tag{9.7}$$

where

$$G_1 = 1 + \frac{1}{2}\phi\alpha^2\sigma^2\sigma_z^2,$$
$$G_2 = 1 - \frac{1}{2}(1+\phi)\alpha^2\sigma^2\sigma_z^2,$$
$$G_3 = 1 - \alpha^2\sigma^2\sigma_z^2.$$

9.2 Search and Illiquidity

We next examine how the search friction impacts the illiquidity measures and the expected return. We perform two related but distinct exercises: compare the search market with the centralized market of Section 3, and vary the measure N of meetings between liquidity demanders and suppliers.

When N decreases, the search process becomes less efficient and trading volume decreases. At the same time, the price in each meeting remains the same because it depends only on the distribution of bargaining power within the meeting. Since illiquidity λ measures the price impact of volume, it increases. One would expect that λ in the search market is higher than in the centralized market because only a fraction of suppliers are involved in bilateral meetings and provide liquidity ($N \leq 1 - \pi$). Proposition 9.3 confirms this result when bargaining power is symmetric ($\phi = 1/2$). The result is also true when suppliers have more bargaining power than demanders ($\phi > 1/2$) because the liquidity shock has then larger price impact. Moreover, the result extends to all values of ϕ when less than half of suppliers are involved in meetings ($N \leq (1 - \pi)/2$).

Proposition 9.3. Price impact λ is

$$\lambda = \frac{\alpha\sigma^2(1 + 2\phi)}{2N}, \tag{9.8}$$

and increases when the measure N of meetings decreases. It is higher than in the centralized market if $\phi + 1/2 \geq N/(1 - \pi)$.

Because the price in the search market is independent of N, so is the price reversal γ. Moreover, γ in the search market is higher than in the centralized market if ϕ is large relative to π.

Proposition 9.4. Price reversal γ is

$$\gamma = \frac{\alpha^2\sigma^4\sigma_z^2(1 + 2\phi)^2}{16}, \tag{9.9}$$

and is independent of the measure N of meetings. It is higher than in the centralized market if $\phi + 1/2 \geq 2\pi$.

When the measure N of meetings decreases, agents are less likely to trade in Period 1. A natural conjecture then is that the illiquidity discount increases and so does the asset's expected return. Proposition 9.5 confirms this conjecture under the sufficient condition $\phi \leq 1/2$. Intuitively, if $\phi \approx 1$, a decrease in the measure of meetings does not affect liquidity demanders because they extract no surplus from a meeting. Since, however, liquidity suppliers become worse off, the risk-neutral probability of being a demander decreases, and the price can increase.[1]

Proposition 9.5. A decrease in the measure N of meetings raises the asset's expected return $E(R)$ if $\phi \leq 1/2$.

9.3 Literature

Early work modeling search frictions in asset markets and their implications for equilibrium prices includes Burdett and O'Hara (1987), Pagano (1989b) and Keim and Madhavan (1996). These papers focus on the market for large blocks of shares (known as the "upstairs" market in the New York Stock Exchange).

Duffie et al. (2002, 2005, 2008) model price formation in asset markets building on the search framework of Diamond (1982), Mortensen (1982) and Pissarides (1985), in which a continuum of agents negotiate prices in bilateral meetings over an infinite horizon and continuous time. Duffie et al. (2002) focus on the repo market, where traders can borrow or lend assets. In a centralized market with no frictions, lenders of positive-supply assets would compete their rent down to zero. Indeed, equilibrium requires that some agents hold the assets, and hence would be willing to lend them as long as they earn any non-zero rent. With search frictions, however, lenders can earn a rent because they can extract some of the borrowers' surplus when bargaining in bilateral meetings. The rent is an additional payoff from holding the assets and raises their price in the spot market.

Duffie et al. (2008) focus on the spot market and assume that the valuation of agents for a risky asset switches over time between high and low. Agents with high valuation who do not own the asset seek to buy it. Conversely, agents with low valuation who own the asset seek to sell it. The equilibrium prices that emerge in the bilateral meetings depend not only on the measures of buyers and sellers, as in a centralized

[1] The illiquidity discount in the search market is higher than in the centralized market if ϕ is large relative to π. This property is the same as for λ and γ, but the calculations are more complicated.

market, but also on their relative bargaining power. Our model yields an extreme version of this result: the price in Period 1 depends only on the bargaining power parameter ϕ and not on the measures $(\pi, 1 - \pi)$ of liquidity demanders and suppliers. An implication of this result is that an increase in search frictions can raise or lower the asset price, with the price decreasing when there are more buyers than sellers. Indeed, with larger frictions, the price responds less to the aggregate demand/supply conditions, and hence decreases when these conditions are favorable to the sellers. Finally, following a positive shock to the measure of sellers, which moves the market away from steady state, prices drop and recover gradually with the drop being larger when frictions increase.

Duffie et al. (2005) introduce market makers who intermediate trade. Market makers differ from other agents, who we term investors, because they can be contacted more easily. If investors are better able to contact each other, then market makers face more competition and post lower bid-ask spreads. Moreover, if investors are heterogeneous in their ability to contact market makers, then market makers post lower spreads for investors with higher such ability. Weill (2007) studies the dynamics of an intermediated search market away from steady state. He shows that following a positive shock to the measure of sellers, market makers build up inventories, which they gradually unload to buyers. Market makers acquire the asset despite having lower valuation for it than other agents because they are more efficient in passing it to the buyers.

Vayanos and Wang (2007) and Weill (2008) extend the analysis to multiple assets, and show that search frictions can generate price discrepancies between assets with identical payoffs. Buyers choose one of two assets to search for, and then can only meet sellers of that asset. In equilibrium, they can locate one asset more easily, and are hence willing to pay a higher price for it. The asset that is easier to locate has a higher number of sellers either because it attracts endogenously high-turnover agents in Vayanos and Wang (2007), or because it is in larger supply in Weill (2008). Note that one-asset models, such as Duffie et al. (2008), yield the opposite prediction that assets in larger supply trade at lower prices.

Vayanos and Weill (2008) show that deviations from the law of one price can arise even under simultaneous search, i.e., buyers can meet sellers of all assets. Key to this result is the presence of short sellers, who borrow an asset in the repo market, then sell it in the spot market, and then buy it back again to unwind the short sale. In equilibrium, short sellers endogenously concentrate in one asset, making it more liquid. That asset trades at a higher price because its superior liquidity is priced by the longs, i.e., the buyers who seek to establish long positions. Moreover, the higher concentration of short-sellers in one asset makes it profitable for longs to lend the asset in the repo market, and further raises its price as in Duffie et al. (2002).

A number of papers relax the assumption that agents can hold zero or one unit of an asset. Garleanu (2009) and Lagos and Rocheteau (2009) show that an increase in search frictions makes agents less willing to change their positions in response to short-run shocks to their valuation for the asset. This is because they are aware that it will take them time to change their positions back should an offsetting shock hit. Since agents become less responsive to shocks in either direction, search frictions have an ambiguous effect on the price, consistent with Duffie et al. (2008). Lagos et al. (2012) study the effects of shocks that move the market away from steady state, and show that the speed of recovery is non-monotonic in search frictions. Afonso and Lagos (2011) study price formation in the interbank market, and determine how the Federal Funds Rate depends on the search frictions and on Federal Reserve policy actions. Pagnotta and Philippon (2012) study the competition between financial exchanges that offer different speeds of trade execution, modeled as contact rates, at different fees.

Search models emphasize the idea that matching buyers and sellers takes time. In their work on participation costs, Grossman and Miller (1988) model a related idea: a liquidity shock experienced by some agents is absorbed first by a small set of market participants and then by a larger set of agents who gradually arrive in the market. The market participants who first absorb the shock act as intermediaries, building up inventories and then unwinding them. Search models provide a natural setting to study the process through which assets are reallocated across agents via the temporal variation in intermediaries' inventories.

This is done, for example, in Weill (2007), where intermediaries are modeled as a special class of agents who can be contacted more easily than others. It is also done in Afonso and Lagos (2011), where agents engage endogenously in intermediation when they meet others with large liquidity shocks: they absorb more than their final share of a shock knowing that they can unload it to others in future bilateral meetings. Duffie and Strulovici (2011) model the process through which new agents slowly become informed about liquidity shocks in one market and bring their capital into that market. Mitchell et al. (2007) and Duffie (2010) emphasize the idea that capital moves slowly across markets in response to profitable investment opportunities.

All papers mentioned so far assume that agents have symmetric information about the asset payoff. If some agents receive private signals, then these can be revealed gradually through the bilateral meetings, as agents learn the information of those they meet and of those that their meeting partners have met in the past. Papers studying the transmission of private information in decentralized markets include Wolinksy (1990), Blouin and Serrano (2001), Duffie and Manso (2007), Duffie et al. (2009), Golosov et al. (2011) and Zhu (2012).

Finally, some papers study portfolio choice under the assumption that agents can trade only after a lag, which could reflect unmodeled search frictions or market breakdowns. For example, Longstaff (2001) restricts trading strategies to be of bounded variation, while Ang et al. (2011) assume that investors can trade only at exogenous random times. Both papers take prices as given and compute the utility loss from infrequent trading. This exercise is in the spirit of the one performed in Constantinides (1986) in the case of transaction costs, but the utility loss is larger in the case of infrequent trading. Longstaff (2009) shows in an equilibrium model that infrequent trading has large effects on asset prices.

10

Conclusion

In this paper we survey the theoretical literature on market liquidity. This large and growing literature traces illiquidity, i.e., the lack of liquidity, to underlying market imperfections. It shows that even simple imperfections can break the clean properties of the perfect-market model and lead to rich but complex behavior. Moreover, this behavior can be sensitive to the particular form of imperfection and the specification of the model. The lack of a unified framework and robust predictions makes it difficult not only to advance our theoretical understanding of illiquidity, but also to provide guidance for empirical work.

In this survey we hope to demonstrate that a framework can be constructed to unify the existing theoretical literature. Our framework nests six main imperfections studied in the literature: participation costs, transaction costs, asymmetric information, imperfect competition, funding constraints, and search. These imperfections map into six different theories of illiquidity. Using our framework, we examine in a systematic manner how the six imperfections affect illiquidity and expected asset returns. We also examine how well different empirical measures of illiquidity capture the underlying imperfections. Needless to say, the framework has a number of limitations, some of which are pointed out in Section 1. But this only suggests that more research is needed; and the limitations of the framework may well point us to new and fruitful directions.

Acknowledgments

We thank Viral Acharya, Bruno Biais, Peter DeMarzo, Thierry Foucault, Mike Gallmeyer, Nicolae Garleanu, Denis Gromb, Peter Kondor, Haitao Li, Albert Menkveld, Anya Obizhaeva, Maureen O'Hara, Anna Pavlova, Matt Spiegel, Vish Viswanathan, Pierre-Olivier Weill, and Kathy Yuan for very helpful comments. Financial support from the Paul Woolley Centre at the LSE is gratefully acknowledged.

References

Acharya, V. and L. Pedersen (2005), 'Asset pricing with liquidity risk'. *Journal of Financial Economics* **77**(2), 375–410.

Acharya, V. and S. Viswanathan (2011), 'Leverage, moral hazard, and liquidity'. *Journal of Finance* **66**, 99–138.

Admati, A. (1985), 'A noisy rational expectations equilibrium for multi-asset securities markets'. *Econometrica* **53**, 629–658.

Admati, A. and P. Pfleiderer (1988), 'A theory of intraday patterns: Volume and price variability'. *Review of Financial Studies* **1**, 3–40.

Admati, A. and P. Pfleiderer (1991), 'Sunshine trading and financial market equilibrium'. *Review of Financial Studies* **4**(3), 443–481.

Adrian, T., E. Etula, and H. S. Shin (2009), 'Risk appetite and exchange rates'. Working paper, Princeton University.

Afonso, G. and R. Lagos (2011), 'Trade dynamics in the market for Federal Funds'. Working paper, New York University.

Aiyagari, R. and M. Gertler (1991), 'Asset returns with transaction costs and uninsurable individual risks: A stage III exercise'. *Journal of Monetary Economics* **27**, 309–331.

Aiyagari, R. and M. Gertler (1999), 'Overreaction of asset prices in general equilibrium'. *Review of Economic Dynamics* **2**, 3–35.

Akerlof, G. A. (1970), 'The market for lemons: Quality uncertainty and the market mechanism'. *The Quaterly Journal of Economics* **84**(3), 359–369.

Albagli, E. (2011), 'Amplification of uncertainty in illiquid markets'. Working paper, University of Southern California.

Allen, F. and D. Gale (1994), 'Limited market participation and volatility of asset prices'. *American Economic Review* **84**(4), 933–955.

Allen, F. and D. Gale (2000), 'Bubbles and crises'. *Economic Journal* **110**, 236–255.

Almgren, R. (2003), 'Optimal execution with nonlinear impact functions and trading-enhanced risk'. *Applied Mathematical Finance* **10**, 1–18.

Almgren, R. and N. Chriss (1999), 'Value under liquidation'. *Risk* **12**, 61–63.

Amihud, Y. and H. Mendelson (1980), 'Dealership market: Market-making with inventory'. *Journal of Financial Economics* **8**, 31–53.

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. Pedersen (2005), 'Liquidity and asset pricing'. *Foundations and Trends in Finance* **1**, 269–364.

Ang, A., D. Papanikolaou, and M. Westerfield (2011), 'Portfolio choice with illiquid assets'. Working paper, Columbia University.

Attari, M., A. Mello, and M. Ruckes (2005), 'Arbitraging arbitrageurs'. *Journal of Finance* **60**(5), 2471–2511.

Back, K. (1992), 'Insider trading in continuous time'. *Review of Financial Studies* **5**(3), 387–409.

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

Back, K. and S. Baruch (2011), 'Strategic liquidity provision in limit order markets'. Working paper, Rice University.

Back, K., C. Cao, and G. Willard (2000), 'Imperfect competition among informed traders'. *Journal of Finance* **55**(5), 2117–2155.

Back, K. and H. Pedersen (1998), 'Long-lived information and intraday patterns'. *Journal of Financial Markets* **1**, 385–402.

Bagehot, W. (1971), 'The Only Game in Town'. *Financial Analysts Journal* **22**, 12–14.

Balduzzi, P. and A. Lynch (1999), 'Transaction costs and predictability: Some utility cost calculations'. *Journal of Financial Economics* **52**, 47–78.

Baruch, S. (2002), 'Insider trading and risk aversion'. *Journal of Financial Markets* **5**, 451–464.

Basak, S. and B. Croitoru (2000), 'Equilibrium mispricing in a capital market with portfolio constraints'. *Review of Financial Studies* **13**, 715–748.

Basak, S. and B. Croitoru (2006), 'On the role of arbitrageurs in rational markets'. *Journal of Financial Economics* **81**, 143–173.

Basak, S. and D. Cuoco (1998), 'An equilibrium model with restricted stock market participation'. *Review of Financial Studies* **11**(2), 309–341.

Beber, A., J. Driessen, and P. Tuijp (2012), 'Pricing liquidity risk with heterogeneous investment horizons'. Working paper, Cass Business School.

Bernanke, B. and M. Gertler (1989), 'Agency costs, net worth, and business fluctuations'. *The American Economic Review* **1**, 14–31.

Bernhardt, D. and E. Hughson (1997), 'Splitting orders'. *Review of Financial Studies* **10**(1), 69–102.

Bertsimas, D. and A. Lo (1998), 'Optimal control of execution costs'. *Journal of Financial Markets* **1**, 1–50.

Bhattacharya, U. and M. Spiegel (1991), 'Insiders, outsiders, and market breakdowns'. *Review of Financial Studies* **4**, 255–282.

Biais, B. (1993), 'Price formation and equilibrium liquidity in fragmented and centralized markets'. *Journal of Finance* **48**, 157–185.

Biais, B., L. Glosten, and C. Spatt (2005), 'Market microstructure: A survey of microfoundations, empirical results and policy implications'. *Journal of Financial Markets* **8**, 217–264.

Biais, B., F. Heider, and M. Hoerova (2012), 'Risk-sharing or Risk-taking? Counterparty risk, incentives and margins'. Working paper, University of Toulouse.

Biais, B., J. Hombert, and P.-O. Weill (2011). Trading and liquidity with limited cognition.

Biais, B., D. Martimort, and J.-C. Rochet (2000), 'Competing mechanisms in a common value environment'. *Econometrica* **68**(4), 799–837.

Blouin, M. and R. Serrano (2001), 'A decentralized market with common values uncertainty: Non-steady states'. *Review of Economic Studies* **68**, 323–346.

Brown, D. and R. Jennings (1990), 'On technical analysis'. *Review of Financial Studies* **2**, 527–552.

Brunnermeier, M. and L. Pedersen (2005), 'Predatory trading'. *Journal of Finance* **60**(4), 1825–1863.

Brunnermeier, M. and L. Pedersen (2009), 'Market liquidity and funding liquidity'. *Review of Financial Studies* **22**, 2201–2238.

Buffa, A. (2011), 'Insider trade disclosure, market efficiency, and liquidity'. Working paper, London Business School.

Burdett, K. and M. O'Hara (1987), 'Building blocks: An introduction to block trading'. *Journal of Banking and Finance* **11**, 193–212.

Buss, A. and B. Dumas (2011), 'The equilibrium dynamics of liquidity and illiquid asset prices'. Working paper, Goethe University Frankfurt.

Buss, A., R. Uppal, and G. Vilkov (2011), 'Asset prices in general equilibrium with transactions costs and recursive utility'. Working paper, Goethe University Frankfurt.

Caldentey, R. and E. Stacchetti (2010), 'Insider trading with a random deadline'. *Econometrica* **78**, 245–283.

Cao, H., M. Evans, and R. Lyons (2006), 'Inventory information'. *Journal of Business* **79**, 325–364.

Carlin, B., M. Lobo, and S. Viswanathan (2007), 'Episodic liquidity crises: Cooperative and predatory trading'. *Journal of Finance* **62**(5), 2235–2274.

Casamatta, C. and S. Pouget (2011), 'Fund managers' contracts and financial markets' short-termism'. Working paper, University of Toulouse.

Cespa, G. and T. Foucault (2011), 'Learning from prices, liquidity spillovers and endogenous market segmentation'. Working paper, HEC Paris.

Cespa, G. and X. Vives (2012), 'Expectations, illiquidity, and short-term trading'. Working paper, City University.

Chabakauri, G. (2012), 'Asset pricing in general equilibrium with constraints'. Working paper, London School of Economics.

Chau, M. and D. Vayanos (2008), 'Strong form efficiency with monopolistic insiders'. *Review of Financial Studies* **21**, 2275–2306.

Chowhdry, B. and V. Nanda (1991), 'Multimarket trading and market liquidity'. *Review of Financial Studies* **4**(3), 483–511.

Cohen, K. J., S. F. Maier, R. A. Schwartz, and D. K. Whitcomb (1981), 'Transaction costs, order placement strategy, and the existence of the bid-ask spread'. *Journal of Political Economy* **89**, 287–305.

Constantinides, G. M. (1986), 'Capital market equilibrium with transaction costs'. *Journal of Political Economy* **94**(4), 842–862.

Copeland, T. and D. Galai (1983), 'Information effects on the bid-ask spread'. *Journal of Finance* **38**, 1457–1469.

Cuoco, D. (1997), 'Optimal consumption and equilibrium prices with portfolio constraints and stochastic income'. *Journal of Economic Theory* **72**(1), 33–73.

Danielsson, J., H. S. Shin, and J.-P. Zigrand (2012), 'Balance sheet capacity and endogenous risk'. Working paper, London School of Economics.

Davis, M. H. A. and A. Norman (1990), 'Portfolio selection with transaction costs'. *Mathematics of Operations Research* **15**, 676–713.

De Long, B., A. Shleifer, L. Summers, and R. Waldmann (1990), 'Noise trader risk in financial markets'. *Journal of Political Economy* **98**, 703–738.

DeMarzo, P. and B. Urosevic (2006), 'Ownership dynamics and asset pricing with a large shareholder'. *Journal of Political Economy* **114**(4), 774–815.

Demsetz, H. (1968), 'The cost of transacting'. *The Quarterly Journal of Economics* **82**, 33–53.

Detemple, J. and S. Murthy (1997), 'Equilibrium asset prices and no-arbitrage with portfolio constraints'. *Review of Financial Studies* **10**(4), 1133–1174.

Diamond, D. and R. Verrecchia (1981), 'Information aggregation in a noisy rational expectations economy'. *Journal of Financial Economics* **9**, 221–235.

Diamond, P. A. (1982), 'Aggregate demand management in search equilibrium'. *Journal of Political Economy* **90**, 881–894.

Dow, J. and G. Gorton (1994), 'Arbitrage chains'. *Journal of Finance* **49**(3), 819–49.

Duffie, D. (2010), 'Presidential address: Asset price dynamics with slow-moving capital'. *Journal of Finance* **65**, 1237–1267.

Duffie, D., N. Garleanu, and L. Pedersen (2002), 'Securities lending, shorting, and Pricing'. *Journal of Financial Economics* **66**, 307–339.

Duffie, D., N. Garleanu, and L. Pedersen (2005), 'Over-the-counter markets'. *Econometrica* **73**, 1815–1847.

Duffie, D., N. Garleanu, and L. Pedersen (2008), 'Valuation in over-the-counter markets'. *Review of Financial Studies* **20**, 1865–1900.

Duffie, D., S. Malamud, and G. Manso (2009), 'Information percolation with equilibrium search dynamics'. *Econometrica* **77**, 1513–1574.

Duffie, D. and G. Manso (2007), 'Information percolation in large markets'. *American Economic Review Papers and Proceedings* **97**, 203–209.

Duffie, D. and B. Strulovici (2011), 'Capital mobility and asset pricing'. Working paper, Stanford University.

Dumas, B. and E. Luciano (1991), 'An exact solution to a dynamic portfolio choice problem under transactions costs'. *Journal of Finance* **46**(2), 577–595.

Easley, D. and M. O'Hara (1987), 'Price, trade size, and information in securities markets'. *Journal of Financial Economics* **19**, 69–90.

Easley, D. and M. O'Hara (1992), 'Time and the process of security price adjustment'. *Journal of Finance* **47**, 576–605.

Easley, D. and M. O'Hara (2004), 'Information and the cost of capital'. *Journal of Finance* **59**, 1553–1583.

Edirisinghe, C., V. Naik, and R. Uppal (1993), 'Optimal replication of options with transaction costs and trading restrictions'. *Journal of Financial and Quantitative Analysis* **28**, 117–138.

Eisfeldt, A. (2004), 'Endogenous liquidity in asset markets'. *Journal of Finance* **59**, 1–30.

Ellul, A. and M. Pagano (2006), 'IPO underpricing and after-market liquidity'. *Review of Financial Studies* **19**, 381–421.

Fardeau, V. (2011), 'Strategic liquidity provision and predatory trading'. Working paper, London School of Economics.

Fishman, M. and K. Hagerty (1992), 'Insider trading and the efficiency of stock prices'. *RAND Journal of Economics* **23**, 106–122.

Fleming, W., S. Grossman, J.-L. Vila, and T. Zariphopoulou (1990), 'Optimal portfolio rebalancing with transactions costs'. Working paper, Brown University.

Foster, D. and S. Viswanathan (1996), 'Strategic trading when agents forecast the forecasts of others'. *Journal of Finance* **51**(4), 1437–1478.

Foucault, T. (1999), 'Order flow composition and trading costs in a dynamic limit order market'. *Journal of Financial Markets* **2**, 99–134.

Foucault, T., O. Kadan, and E. Kandel (2005), 'Limit order book as a market for liquidity'. *Review of Financial Studies* **18**(4), 1171–1217.

Garleanu, N. (2009), 'Portfolio choice and pricing in illiquid markets'. *Journal of Economic Theory* **144**, 532–564.

Garleanu, N. and L. Pedersen (2004), 'Adverse selection and the required return'. *Review of Financial Studies* **17**(3), 643–665.

Garleanu, N. and L. Pedersen (2011), 'Margin-based asset pricing and deviations from the law of one price'. *Review of Financial Studies* **24**(6), 1980–2022.

Garman, M. (1976), 'Market microstructure'. *Journal of Financial Economics* **3**, 257–275.

Geanakoplos, J. (1997), 'Promises, promises'. In: B. Arthur, S. Durlauf, and D. Lane (eds.): *The Economy as an Evolving Complex System II*. Reading, MA: Addison-Wesley, pp. 285–320.

Geanakoplos, J. (2003), 'Liquidity, default and crashes: Endogenous contracts in general equilibrium'. In: M. Dewatripont, L. Hansen, and S. Turnovsky (eds.): *Advances in Economics and Econometrics: Theory and Applications II, Econometric Society Monographs: Eighth World Congress*. Cambridge, UK: Cambridge University Press, pp. 170–205.

Geanakoplos, J. and A. Fostel (2008), 'Leverage cycles and the anxious economy'. *American Economic Review* **98**, 1211–1244.

Glosten, L. (1989), 'Insider trading, liquidity and the role of the monopolist specialist'. *Journal of Business* **62**(2), 211–235.

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

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

Goettler, R., C. Parlour, and U. Rajan (2005), 'Equilibrium in a dynamic limit order market'. *Journal of Finance* **60**, 2149–2192.

Goldsmith, D. (1976), 'Transaction costs and the theory of portfolio selection'. *Journal of Finance* **31**, 1127–1139.

Golosov, M., G. Lorenzoni, and A. Tsyvinski (2011), 'Decentralized trading with private information'. Working paper, Massachusetts Institute of Technology.

Gromb, D. and D. Vayanos (2002), 'Equilibrium and welfare in markets with financially constrained arbitrageurs'. *Journal of Financial Economics* **66**, 361–407.

Gromb, D. and D. Vayanos (2010), 'Limits of arbitrage'. *Annual Review of Financial Economics* **2**, 251–275.

Gromb, D. and D. Vayanos (2011a), 'The dynamics of financially constrained arbitrage'. Working paper, INSEAD.

Gromb, D. and D. Vayanos (2011b), 'Financially constrained arbitrage and cross-market contagion'. Working paper, INSEAD.

Grossman, S. (1976), 'On the efficiency of competitive stock markets when traders have diverse information'. *Journal of Finance* **31**, 573–585.

Grossman, S. and M. Miller (1988), 'Liquidity and market structure'. *Journal of Finance* **43**, 617–637.

Grossman, S. and J. Stiglitz (1980), 'On the impossibility of informationnaly efficient markets'. *American Economic Review* **70**(3), 393–408.

Grossman, S. and J.-L. Vila (1992), 'Optimal investment strategies with leverage constraints'. *Journal of Financial and Quantitative Analysis* **27**, 151–168.

Grundy, B. and M. McNichols (1989), 'Trade and the revelation of information through prices and direct disclosure'. *Review of Financial Studies* **2**, 495–526.

Guo, M. and H. Ou-Yang (2010), 'A continuous-time model of risk-averse strategic trading with dynamic information'. Working paper, Chueng-Kong Graduate School of Business.

Hart, O. and J. Moore (1994), 'A theory of debt based on the inalienability of human capital'. *Quarterly Journal of Economics* **101**, 841–879.

Hart, O. and J. Moore (1995), 'An analysis of the role of hard claims in constraining management'. *American Economic Review* **85**, 567–585.

Hasbrouck, J. (2007), *Empirical Market Microstructure*. Oxford: Oxford University Press.

He, H. and J. Wang (1995), 'Differential information and dynamic behavior of stock trading volume'. *Review of Financial Studies* **8**(4), 919–972.

He, Z. and A. Krishnamurthy (2012), 'A model of capital and crises'. *Review of Economic Studies* **79**, 735–777.

Heaton, J. and D. J. Lucas (1996), 'Evaluating the effects of incomplete markets on risk sharing and asset pricing'. *Journal of Political Economy* **104**, 443–487.

Hellwig, M. (1980), 'On the aggregation of information in competitive markets'. *Journal of Economic Theory* **22**, 477–498.

Hirshleifer, J. (1971), 'The private and social value of information and the reward to inventive activity'. *American Economic Review* **61**, 561–574.

Ho, T. and H. Stoll (1980), 'On dealer markets under competition'. *Journal of Finance* **35**(2), 259–267.

Ho, T. and H. Stoll (1981), 'Optimal dealer pricing under trading transactions and return uncertainty'. *Journal of Financial Economics* **9**, 47–73.

Holden, C. and A. Subrahmanyam (1992), 'Long-lived private information and imperfect competition'. *Journal of Finance* **47**(1), 247–270.

Holden, C. and A. Subrahmanyam (1994), 'Risk aversion, imperfect competition, and long-lived information'. *Economic Letters* **44**, 181–190.

Hombert, J. and D. Thesmar (2011), 'Overcoming limits of arbitrage: Theory and evidence'. Working paper, HEC.

Huang, J. and J. Wang (2009), 'Liquidity and market crashes'. *Review of Financial Studies* **22**, 2607–1643.

Huang, J. and J. Wang (2010), 'Market liquidity, asset prices, and welfare'. *Journal of Financial Economics* **95**, 107–127.

Huang, M. (2003), 'Liquidity shocks and equilibrium liquidity premia'. *Journal of Economic Theory* **109**, 104–129.

Huberman, G. and W. Stanzl (2005), 'Optimal liquidity trading'. *Review of Finance* **9**, 165–200.

Huddart, S., J. Hughes, and C. Levine (2001), 'Public disclosure and dissimulation of insider trades'. *Econometrica* **69**(3), 665–681.

Jackson, M. (1991), 'Equilibrium, price formation, and the value of private information'. *Review of Financial Studies* **4**(1), 1–16.

Jang, B.-G., H. K. Koo, H. Liu, and M. Loewenstein (2007), 'Liquidity premia and transaction costs'. *Journal of Finance* **62**, 2329–2366.

Jouini, E. and H. Kallal (1995), 'Martingales and arbitrage in securities markets with transaction costs'. *Journal of Economic Theory* **66**, 178–197.

Jurek, J. and H. Yang (2007), 'Dynamic portfolio selection in arbitrage'. Working paper, Princeton University.

Keim, D. and A. Madhavan (1996), 'The upstairs market for large-block transactions: Analysis and measurement of price effects'. *Review of Financial Studies* **9**, 1–36.

Kiyotaki, N. and J. Moore (1997), 'Credit Cycles'. *Journal of Political Economy* **105**, 211–248.

Klemperer, P. and M. Meyer (1989), 'Supply function equilibria in oligopoly under uncertainty'. *Econometrica* **57**, 1243–1277.

Kondor, P. (2009), 'Risk in dynamic arbitrage: Price effects of convergence trading'. *Journal of Finance* **64**, 638–658.

Kyle, A. (1985), 'Continuous auctions and insider trading'. *Econometrica* **53**(6), 1315–1336.

Kyle, A. (1989), 'Informed speculation with imperfect competition'. *Review of Economic Studies* **56**, 317–356.

Kyle, A. and W. Xiong (2001), 'Contagion as a wealth effect'. *Journal of Finance* **56**, 1401–1440.

Laffont, J.-J. and E. Maskin (1990), 'The efficient market hypothesis and insider trading on the stock market'. *Journal of Political Economy* **98**, 70–93.

Lagos, R. and G. Rocheteau (2009), 'Liquidity in asset markets with search frictions'. *Econometrica* **77**, 403–426.

Lagos, R., G. Rocheteau, and P.-O. Weill (2012), 'Crises and liquidity in over the counter markets'. *Journal of Economic Theory* **146**, 2169–2205.

Leland, H. (1992), 'Insider trading: Should it be prohibited'. *Journal of Political Economy* **100**, 859–887.

Leland, H. and M. Rubinstein (1985), 'Option pricing and replication with transaction costs'. *Journal of Finance* **40**(5), 1283–1301.

Liu, H. (2004), 'Optimal consumption and investment with transaction costs and multiple risky assets'. *Journal of Finance* **59**(1), 289–338.

Liu, H. and M. Loewenstein (2002), 'Optimal portfolio selection with transaction costs and finite horizons'. *Review of Financial Studies* **15**, 805–835.

Liu, H. and Y. Wang (2012), 'Over-the-counter markets: Market making with asymmetric information, inventory risk and imperfect competition'. Working paper, Washington University.

Liu, J. and F. Longstaff (2004), 'Losing money on arbitrage: Optimal dynamic portfolio choice in markets with arbitrage opportunities'. *Review of Financial Studies* **17**(3), 611–641.

Lo, A., H. Mamaysky, and J. Wang (2004), 'Asset prices and trading volume under fixed transactions costs'. *Journal of Political Economy* **112**, 1054–1090.

Longstaff, F. (2001), 'Optimal portfolio choice and the valuation of illiquid securities'. *Review of Financial Studies* **14**(2), 407–431.

Longstaff, F. (2009), 'Portfolio claustrophobia: Asset pricing in markets with illiquid assets'. *American Economic Review* **99**, 1119–1144.

Luttmer, E. G. (1996), 'Asset pricing in economies with frictions'. *Econometrica* **64**, 1439–1467.

Lynch, A. and P. Balduzzi (2000), 'Predictability and transaction costs: The impact on rebalancing rules and behavior'. *Journal of Finance* **55**, 2285–2310.

Lynch, A. and S. Tan (2011), 'Explaining the magnitude of liquidity premia: The role of return predictability, wealth shocks and state-dependent transaction costs'. *Journal of Finance* **66**, 1329–1368.

Madhavan, A. (2000), 'Market microstructure: A survey'. *Journal of Financial Markets* **3**, 205–258.

Mankiw, N. and S. Zeldes (1991), 'The consumption of stockholders and nonstockholders'. *Journal of Financial Economics* **29**, 97–112.

Mayshar, J. (1979), 'Transaction costs in a model of capital market equilibrium'. *Journal of Political Economy* **87**, 673–700.

Mehra, R. and E. C. Prescott (1985), 'The equity premium: A puzzle'. *Journal of Monetary Economics* **15**(2), 145–161.

Merton, R. (1971), 'Optimum consumption and portfolio rules in a continuous-time model'. *Journal of Economic Theory* **3**, 373–413.

Merton, R. (1987), 'Presidential address: A simple model of capital market equilibrium with incomplete information'. *Journal of Finance* **42**, 483–510.

Milbradt, K. (2012), 'Level 3 assets: Booking profits, concealing losses'. *Review of Financial Studies* **25**, 55–95.

Mildenstein, E. and H. Schleef (1983), 'The optimal pricing policy of a monopolistic marketmaker in equity market'. *Journal of Finance* **38**, 218–231.

Mitchell, M., L. Pedersen, and T. Pulvino (2007), 'Slow moving capital'. *American Economic Review, Papers and Proceedings* **97**, 215–220.

Mortensen, D. (1982), 'Property rights and efficiency in mating, racing, and related games'. *American Economic Review* **72**, 968–979.

Naik, N., A. Neuberger, and S. Viswanathan (1999), 'Trade disclosure regulation in markets with negotiated trade'. *Review of Financial Studies* **12**, 873–900.

Obizhaeva, A. and J. Wang (2006), 'Optimal trading strategy and supply/demand dynamics'. Working paper, University of Maryland.

O'Hara, M. (1995), *Market Microstructure Theory*. Cambridge: Blackwell Publishers.

O'Hara, M. (2003), 'Liquidity and price discovery'. *Journal of Finance* **58**, 1335–1354.

Pagano, M. (1989a), 'Endogenous market thinness and stock price volatility'. *Review of Economic Studies* **56**, 269–287.

Pagano, M. (1989b), 'Trading volume and asset liquidity'. *Quarterly Journal of Economics* **104**, 255–274.

Pagano, M. and A. Roell (1996), 'Transparency and liquidity: A comparison of auction and dealer markets with informed trading'. *Journal of Finance* **51**, 579–611.

Pagnotta, E. and T. Philippon (2012), 'Competing on speed'. Working paper, New York University.

Parlour, C. (1998), 'Price dynamics in limit order markets'. *Review of Financial Studies* **1**, 789–816.

Parlour, C. and D. Seppi (2008), 'Limit order markets: A survey'. In: A. Boot and A. Thakor (eds.): *Handbook of Financial Intermediation and Banking*. North Holland.

Pavlova, A. and R. Rigobon (2008), 'The role of portfolio constraints in the international propagation of shocks'. *Review of Economic Studies* **75**, 1215–1256.

Pissarides, C. (1985), 'Short-run equilibrium dynamics of unemployment, vacancies, and real wages'. *American Economic Review* **75**, 676–690.

Pritsker, M. (2005), 'Large investors: Implications for equilibrium asset, returns, shock absorption, and liquidity'. Working paper, Board of Governors of the Federal Reserve.

Qiu, W. and J. Wang (2010), 'Asset pricing under heterogeneous information'. Working Paper, Massachusetts Institute of Technology.

Repullo, R. (1999), 'Some remarks on Leland's model of insider trading'. *Economica* **66**(263), 359–374.

Rostek, M. and M. Weretka (2011), 'Dynamic thin markets'. Working paper, University of Wisconsin.

Rosu, I. (2009), 'A dynamic model of the limit-order book'. *Review of Financial Studies* **22**, 4601–4641.

Rytchkov, O. (2011), 'Asset pricing with dynamic margin constraints'. Working paper, Temple University.

Shleifer, A. and R. Vishny (1992), 'Liquidation values and debt capacity: A market equilibrium approach'. *Journal of Finance* **47**, 1343–1366.

Shleifer, A. and R. Vishny (1997), 'The limits of arbitrage'. *Journal of Finance* **52**, 35–55.

Soner, M., S. Shreve, and J. Cvitanic (1995), 'There is no nontrivial hedging portfolio for option pricing with transaction costs'. *Annals of Applied Probability* **5**, 327–355.

Stoll, H. (1978), 'The supply of dealer services in securities markets'. *Journal of Finance* **33**, 1133–1151.

Suominen, M. and K. Rinne (2011), 'A structural model of short-term reversals'. Working paper, Aalto University.

Tuckman, B. and J.-L. Vila (1992), 'Arbitrage with holding costs: A utility-based approach'. *Journal of Finance* **47**, 1283–1302.

Tuckman, B. and J.-L. Vila (1993), 'Holding costs and equilibrium arbitrage'. Working paper 1153, Anderson Graduate School of Management, UCLA.

Vayanos, D. (1998), 'Transaction costs and asset prices: A dynamic equilibrium model'. *Review of Financial Studies* **11**, 1–58.

Vayanos, D. (1999), 'Strategic trading and welfare in a dynamic market'. *Review of Economic Studies* **66**, 219–254.

Vayanos, D. (2001), 'Strategic trading in a dynamic noisy market'. *Journal of Finance* **56**, 131–171.

Vayanos, D. (2004), 'Flight to quality, flight to liquidity, and the pricing of risk'. Working paper, London School of Economics.

Vayanos, D. and J.-L. Vila (1999), 'Equilibrium interest rate and liquidity premium with transaction costs'. *Economic Theory* **13**, 509–539.

Vayanos, D. and J. Wang (2012a), 'Liquidity and expected returns under asymmetric information and imperfect competition'. *Review of Financial Studies* **25**, 1339–1365.

Vayanos, D. and J. Wang (2012b), 'Market liquidity — theory and empirical evidence'. In: G. Constantinides, M. Harris, and R. Stulz (eds.): *Handbook of the Economics of Finance*. North Holland, Amsterdam. forthcoming.

Vayanos, D. and T. Wang (2007), 'Search and endogenous concentration of liquidity in asset markets'. *Journal of Economic Theory* **136**(1), 66–104.

Vayanos, D. and P.-O. Weill (2008), 'A search-based theory of the on-the-run phenomenon'. *Journal of Finance* **63**, 1361–1398.

Venter, G. (2011), 'Financially constrained strategic arbitrage'. Working paper, Copenhagen Business School.

Vives, X. (1995), 'The speed of information revelation in a financial market mechanism'. *Journal of Economic Theory* **67**, 178–204.

Wang, J. (1993), 'A model of intertemporal asset prices under asymmetric information'. *Review of Economics Studies* **60**, 249–282.

Wang, J. (1994), 'A model of competitive stock trading volume'. *Journal of Political Economy* **102**, 127–168.

Weill, P.-O. (2007), 'Leaning against the wind'. *Review of Economic Studies* **74**, 1329–1354.

Weill, P.-O. (2008), 'Liquidity premia in dynamic bargaining markets'. *Journal of Economic Theory* **140**, 66–96.

Wilson, R. (1979), 'Auctions of shares'. *Quarterly Journal of Economics* **93**, 675–689.

Wolinksy, A. (1990), 'Information revelation in a market with pairwise meetings'. *Econometrica* **58**, 1–23.

Xiong, W. (2001), 'Convergence trading with wealth effects: An amplification mechanism in financial markets'. *Journal of Financial Economics* **62**, 247–292.

Yuan, K. (2005), 'Asymmetric price movements and borrowing constraints: A REE model of crisis, contagion, and confusion'. *Journal of Finance* **60**(1), 379–411.

Zhu, H. (2012), 'Finding a good price in opaque over-the-counter markets'. *Review of Financial Studies* **25**(4), 1255–1285.