ELSEVIER

Journal of Financial Markets 7 (2004) 271–299

Journal of FINANCIAL MARKETS

www.elsevier.com/locate/econbase

Market liquidity as a sentiment indicator ☆

Malcolm Baker[a,*], Jeremy C. Stein[b,c]

[a] Harvard Business School, Harvard University, Boston, MA 02163, USA
[b] Harvard Economics Department, Harvard University, Cambridge, MA 02138, USA
[c] National Bureau of Economic Research, Cambridge, MA 02138, USA

Received 1 October 2003; accepted 24 November 2003

Abstract

We build a model that helps to explain why increases in liquidity—such as lower bid–ask spreads, a lower price impact of trade, or higher turnover–predict lower subsequent returns in both firm-level and aggregate data. The model features a class of irrational investors, who underreact to the information contained in order flow, thereby boosting liquidity. In the presence of short-sales constraints, high liquidity is a symptom of the fact that the market is dominated by these irrational investors, and hence is overvalued. This theory can also explain how managers might successfully time the market for seasoned equity offerings, by simply following a rule of thumb that involves issuing when the SEO market is particularly liquid. Empirically, we find that: (i) aggregate measures of equity issuance and share turnover are highly correlated; yet (ii) in a multiple regression, both have incremental predictive power for future equal-weighted market returns.

JEL classification: G12; G14

Keywords: Liquidity; Sentiment; Short-sales constraints

☆ We thank Michael Brennan, Josh Coval, Ken French, Anthony Lynch, Andrei Shleifer, Avanidhar Subrahmanyam, Tuomo Vuolteenaho and seminar participants at the 2002 American Finance Association Annual Meetings, the Atlanta Finance Forum (Georgia Tech, Georgia State, and Emory Universities), Harvard Business School, the NBER Behavioral Finance Conference, and the Texas Finance Festival for helpful comments. Financial support from the Division of Research of the Harvard Graduate School of Business Administration (Baker) and from the National Science Foundation (Stein) is gratefully acknowledged.

*Corresponding author.
E-mail address: mbaker@hbs.edu (M. Baker).

1386-4181/$ - see front matter
doi:10.1016/j.finmar.2003.11.005

1. Introduction

A growing body of empirical evidence suggests that liquidity predicts stock returns, both at the firm level and in the time series of the aggregate market. Amihud and Mendelson (1986), Brennan and Subrahmanyam (1996), and Brennan et al. (1998) find that measures of increased liquidity, including a low price impact of trade, low bid–ask spreads and high share turnover, are associated with lower future returns in cross sections of individual firms. More recently, Chordia et al. (2000, 2001), Hasbrouck and Seppi (2001), and Huberman and Halka (2001) document that there is considerable time-variation in market-wide liquidity, and Amihud (2002) and Jones (2002) show that these market-wide movements in liquidity also forecast aggregate returns.[1]

The traditional explanation for why liquidity might affect expected returns is a straightforward one (Amihud and Mendelson, 1986; Vayanos, 1998). Investors anticipate having to sell their shares at some point in the future, and recognize that when they do so, they will face transactions costs. These costs can stem either from the inventory considerations of risk-averse market makers or from problems of adverse selection.[2] But in either case, when the transactions costs are greater, investors rationally discount the asset in question by more. This story would seem to fit most naturally with the purely cross-sectional results. In particular, if we compare two stocks, and one is observed to have permanently lower bid–ask spreads and price impacts than the other, as well as higher turnover, it is plausible that the more liquid stock would have a somewhat higher price, and hence lower expected returns.

It is less clear whether the same story can be carried over without modification to explain the time-series results for the aggregate market. First of all, we do not have a well-developed understanding of what drives the common time-series variation in measures of liquidity. For example, though it is a possibility, it seems more of a stretch to argue that there are large swings in the degree of asymmetric information about the market as a whole. Second, as Jones (2002) shows, and as we verify below, the predictive power of aggregate liquidity for market returns, particularly for equal-weighted returns, is large. In a univariate regression, a one-standard-deviation increase in stochastically detrended turnover (equivalent to turnover going from, say, the 1932–1998 mean of 30 percent up to 42 percent in a given year) reduces expected returns on the CRSP equal-weighted index over the next year by approximately 13 percent.

In this paper, we develop an alternative theory to explain the connection between liquidity and expected returns.[3] Our focus is on understanding why *time-variation* in

[1] Somewhat more subtly, Pastor and Stambaugh (2003) argue that expected returns are higher on stocks with a greater sensitivity to market-wide liquidity—i.e., that aggregate liquidity is a priced risk factor.

[2] On the former, see Demsetz (1968), Garman (1976), Stoll (1978), Amihud and Mendelson (1980), and Grossman and Miller (1988). On the latter, see Copeland and Galai (1983), Glosten and Milgrom (1985), Kyle (1985), Easley and O'Hara (1987) and Admati and Pfleiderer (1988).

[3] Although our focus is on the stock market, the link between high prices and market liquidity seems to be pervasive. See, e.g., Shleifer and Vishny (1992) and Stein (1995) for models of the market for corporate asset sales and the housing market, respectively. We discuss the relationship of our theory to this work below.

liquidity, either at the firm level or for the market as a whole, might forecast changes in returns. We implicitly accept the premise that the traditional theory is best suited to explaining why *permanent cross-firm differences* in liquidity are associated with permanent cross-firm differences in expected returns.[4]

Our model rests on two sets of assumptions—one about market frictions, and the other about investor behavior. With respect to the former, we assume that there are short-sales constraints. With respect to the latter, we posit the existence of a class of irrationally overconfident investors, where we think of overconfidence as a tendency to overestimate the relative precision of one's own private signals. In our setting, this form of overconfidence has two distinct manifestations. First, when overconfident investors receive private signals, they tend to overweight them; this leads to ''sentiment shocks'' that can be either positive or negative. Second, when overconfident investors observe the trading decisions of others, they tend to underreact to the information contained in these decisions, since they (erroneously) consider others to be less well-informed than they are. This aspect of overconfidence lowers the price impact of trades, thus boosting liquidity generally.[5]

Given these assumptions, our story goes as follows. At some initial date, the irrational investors receive private signals about future fundamentals, which they overreact to, generating sentiment shocks. The short-sales constraint implies that irrational investors will only be active in the market when their valuations are higher than those of rational investors—i.e., when their sentiment is positive and when the market is, as a result, overvalued. When the sentiment of irrational investors is negative, the short-sales constraint keeps them out of the market altogether. At a subsequent date, there is a round of trading by an informed insider. Since the irrational investors also tend to make the market more liquid in the face of such informed trading, measures of liquidity provide an indicator of the relative presence or absence of these investors, and hence of the level of prices relative to fundamentals.

This theory also provides a novel perspective on a set of issues in corporate finance which have been the focus of much work recently. Stigler (1964), Ritter (1991), Loughran and Ritter (1995), Speiss and Affleck-Graves (1995), and Brav and Gompers (1997), among others, find that firms that issue equity have low stock returns in the subsequent few years—this is the so-called ''new issues puzzle''. Baker and Wurgler (2000) uncover an analogous pattern in the aggregate data: if economy-wide equity issuance is high in a given year, the market as a whole performs poorly in the next year. The usual interpretation of these facts is that the managers making issuance decisions are ''smart money'': they have a better estimate of the long-run fundamental value of their firms than is embodied in the current market price, and they purposefully time their financing decisions to exploit this advantage.[6]

[4] Note that if one measures the cross-sectional link between liquidity and expected returns with Fama–MacBeth (1973) regressions—as is common in the literature—this will pick up any effects coming from either transient or permanent firm-level differences in liquidity.

[5] Odean (1998a) and Kyle and Wang (1997) use a similar mechanism to tie overconfidence to liquidity. But these models make no predictions about the relationship between liquidity and expected returns.

[6] See Stein (1996) for a model along these lines.

We do not dispute that this smart-money mechanism may be part of what is going on. After all, in Graham and Harvey (2001), managers place market timing high on their list of reasons to issue equity. However, our model offers a potentially complementary way of rationalizing these phenomena, without requiring a high degree of managerial timing ability. Whether or not managers make an attempt—smart or misguided—to come up with independent estimates of fundamental value, their financing decisions may still convey information about future returns, if they follow a simple and plausible rule of thumb. In particular, suppose that managers are more willing to issue equity in periods when the market for new offerings is more liquid, in the sense of there being a reduced adverse price impact upon the announcement of a new issue.[7] If they behave this way, their financing choices will be a passive mirror of market liquidity, and will thus, for the reasons outlined above, tend to forecast returns. Again, this mechanism can work even if managers never bother to take a stand on the relationship between prices and long-run fundamental value.

We view the contribution of this paper to be primarily a theoretical one, and as such do not attempt to provide a definitive empirical test of the model. Nevertheless, we do briefly examine some aggregate data on turnover, equity issuance and stock returns, and document the following patterns. First, consistent with the corporate-finance element of our theory, there is a very strong correlation between turnover in a given year and the share of equity in total external finance. The simple correlation coefficient between the two variables is as high as 0.64 (in the period prior to the deregulation of the brokerage industry), and the strength of this relationship is largely unaffected by standard controls for valuation levels, such as the dividend-price ratio, and past returns. Thus our premise that equity issuance is a mirror of market liquidity seems to be borne out in the data.

Second, both turnover and the equity share have considerable forecasting power for year-ahead returns, especially when we focus on an equal-weighted, as opposed to a value-weighted index. This is true when each variable is considered separately from the other; in this respect we are just confirming the earlier work of Jones (2002) and Baker and Wurgler (2000). Moreover, in spite of their high correlation with one another, each plays a significant role when they are entered in the regressions together, and the overall explanatory power for future returns is substantially augmented. In the context of our model, this can be thought of as reflecting the notion that both turnover and the equity share are noisy measures of "true" market liquidity.

The third message that we take away from our brief empirical exercise is that the forecasting power of turnover appears to be large in economic terms. As already noted, in a simple univariate regression, a one-standard-deviation increase in

[7] As we discuss in more detail below, one way to think about the objective function underlying this rule of thumb is that managers care about market liquidity per se—i.e., they simply wish to avoid large price impacts when issuing equity. Alternatively, they may understand that high liquidity is a signal of overvaluation, and may act as if they care about it for this reason. Although either interpretation works equally well, we stress the former, because we want to make the point that managers may appear to time the market successfully even if they have no direct intention of doing so.

detrended turnover implies a downward revision in year-ahead equal-weighted expected returns of roughly 13 percent. While we do not have a specific calibration of the effects that might be generated by a more traditional model, and while the standard error associated with our point estimate is large, this estimate would appear to cast doubt on the notion that the time-variation in expected market returns arises solely from the reaction of rational investors to fluctuations in trading costs.

The rest of the paper proceeds as follows. In Section 2, we develop our basic model, which shows how measures of secondary-market liquidity such as price impact and turnover can forecast returns. In Section 3, we extend the model to incorporate firms' equity issuance decisions, and demonstrate how these too can forecast returns. In Section 4, we discuss some of the model's implications in light of existing evidence, and in Section 5, we present our own empirical results. Section 6 concludes.

2. The basic model: investor sentiment and market liquidity

2.1. Assumptions

We model the pricing of a single stock, which is available in supply Q. There are three dates. At time 3, the stock pays a terminal dividend of $F + \eta + \varepsilon$, where ε and η are independent normally distributed shocks that are not made public prior to liquidation. The variance of ε is standardized to unity. The variance of η is assumed to be infinitesimally small—what matters is that it is small relative to the variance of ε. As will become clear, this is just an expositional trick that simplifies the analysis slightly, by keeping the fundamental risk of the stock—and hence the risk premium—constant from time 1 to time 2.

At time 2, there is an "insider" who obtains early private information about the value of η, and who may trade in infinitesimally small quantities based on this private information.[8] Such trades will partially reveal η to outside investors, and we denote by η^{E} the time-2 rational expectation of η based on the information set available to outsiders. We will say more about insider trading behavior and the inference process that determines η^{E} momentarily.[9]

[8] By making the insider's trades at time 2 small, we keep the overall supply of shares in the hands of outsiders at approximately Q, which again simplifies the exposition by keeping the risk premium constant from time 1 to time 2.

[9] In our setup, the insider's private information η is not publicly revealed until liquidation. An alternative approach is to think of η as relatively short-term private information, so that while it is made public at time 3, there is a chance that the liquidating dividend is not paid out until some later time 4. It is straightforward to extend the model in this direction. However, we need a small probability that the liquidating dividend arrives at the same time as η is made public at time 3. Intuitively, we require that a rational investor who believes the stock to be overpriced relative to long-run fundamentals not wish to take a long position at time 2, even if the market is underreacting to positive information about η at this time. As long as there is some chance that the price will converge to fundamental value by time 3, this condition is satisfied and our basic results go through.

In addition to the insider who appears at time 2, there are two types of outside investors who are present at all times. Both types research the stock and formulate estimates of the terminal dividend. Those investors in the first class are "smart" and have rational expectations, so their resulting time-1 estimate of the dividend, which we denote by V_1^S, is simply F. Those investors in the second class are "dumb", and their time-1 estimate of the dividend, V_1^D, can be either greater than or less than F. We let $\delta = (V_1^D - F)$ denote the dumb investors' initial "sentiment," or misvaluation. As noted in the Introduction, one way to motivate the sentiment δ is to think of the dumb investors being overconfident, and overreacting to a noisy private signal.

At time 2, when the insider trades, smart investors make a rational inference about the implications of this trade, and incorporate it fully into their estimates of the terminal dividend. That is, $V_2^S = V_1^S + \eta^E = F + \eta^E$. In contrast, dumb investors underreact to the information embodied in time-2 trading activity. As a simple way of capturing this, we assume that $V_2^D = V_1^D + \theta\eta^E = F + \delta + \theta\eta^E$, where $\frac{1}{2} < \theta < 1$. In other words, dumb investors update their valuations in the right direction, but not far enough.[10]

In principle, there are a number of underlying behavioral mechanisms that might give rise to this sort of underreaction. For example, dumb investors might suffer from a conservatism bias (Edwards, 1968). Or more prosaically, they may simply not be paying attention, and hence may be unaware of the fact that anything newsworthy has happened at time 2. Our preferred interpretation is that because dumb investors are overconfident in the relative precision of their own private information (i.e., their time-1 signal V_1^D), they do not fully appreciate the significance of the insider's information. Again, the appeal of this interpretation is that it nests the two key aspects of dumb investors' behavior—the time-1 sentiment shock and the time-2 underreaction to trading activity—into a single primitive assumption about overconfidence.[11]

Both types of outside investors have constant-absolute-risk-aversion (CARA) utility. The aggregate risk tolerance of the smart group is given by γ^S, while the aggregate risk tolerance of the dumb group is given by γ^D. Both groups are assumed to be subject to short-sales constraints. Thus, at time 2, one period before liquidation, the demand of the smart group, D_2^S, is given by

$$D_2^S = \max\{\gamma^S(V_2^S - P_2), 0\}, \tag{1}$$

where P_2 is the price of the stock at time 2. Similarly, the time-2 demand of the dumb group, D_2^D, is given by

$$D_2^D = \max\{\gamma^D(V_2^D - P_2), 0\}. \tag{2}$$

[10] As will become clear, the requirement that θ exceed $\frac{1}{2}$ is a technical condition that ensures that our version of Kyle's (1985) model has an interior equilibrium solution for the degree of market liquidity.

[11] In any case, it is becoming increasingly clear that a variety of stock-market phenomena can be understood by appealing to investor underreaction of the sort we model. Barberis et al. (1998) and Hong and Stein (1999) argue that patterns like post-earnings announcement drift (Bernard and Thomas, 1989, 1990) and medium-term momentum (Jegadeesh and Titman, 1993) reflect underreaction to news. Klibanoff et al. (1998) and Hong et al. (2000) present further evidence consistent with the underreaction hypothesis.

There has been a renewed appreciation of the potential relevance of short-sales constraints in recent work.[12] In part, this reflects a growing understanding that such constraints arise not only from the direct transactions costs associated with shorting, but also from a variety of institutional frictions, such as the widespread tendency for mutual-fund charters simply to prohibit the taking of short positions (Almazan et al., 2003). While the existence of these kinds of frictions makes it plausible that both types of outside investors in our model might behave in a short-sales-constrained fashion, our key predictions actually only require one of the two types to be constrained. For example, we could equally well have the dumb group—call them retail investors—be constrained and the smart group—call them arbitrageurs—be unconstrained.

2.2. Liquidity, trading volume and expected returns

Given the demand curves in (1) and (2), it is easy to solve for P_2 as a function of smart and dumb investors' time-2 valuations, V_2^{S} and V_2^{D}. Moreover, once P_2 has been pinned down, it follows immediately that $P_1 = \mathrm{E}_1(P_2)$. That is, P_1 is just the rational expectation of P_2 based on information available at time 1, which is obtained simply by taking P_2 and replacing η^{E} with its ex-ante expectation of zero. This is an arbitrage relationship, because both smart and dumb traders share the same time-1 forecast of P_2, and because the one-period-ahead conditional variance of P_2 is negligible—the only news that hits the market at time 2 is news about η, which has infinitesimally small variance. Proposition 1 and Fig. 1 summarize the results for prices.

Proposition 1. *At $t = \{1, 2\}$, prices can be described by their behavior in three distinct regions of investor sentiment.*

Region 1: Low investor sentiment, $V_t^{\mathrm{D}} < V_t^{\mathrm{S}} - (Q/\gamma^{\mathrm{S}})$. In this region, only smart investors participate in the market at time 2, and dumb investors sit out. Prices are given by $P_t = V_t^{\mathrm{S}} - (Q/\gamma^{\mathrm{S}})$.

Region 2: Intermediate investor sentiment, $V_t^{\mathrm{S}} - (Q/\gamma^{\mathrm{S}}) \leqslant V_t^{\mathrm{D}} \leqslant V_t^{\mathrm{S}} + (Q/\gamma^{\mathrm{D}})$. In this region, both groups of investors participate in the market at time 2. Prices are given by $P_t = [\gamma^{\mathrm{S}}/(\gamma^{\mathrm{S}} + \gamma^{\mathrm{D}})]V_t^{\mathrm{S}} + [\gamma^{\mathrm{D}}/(\gamma^{\mathrm{S}} + \gamma^{\mathrm{D}})]V_t^{\mathrm{D}} - [Q/(\gamma^{\mathrm{S}} + \gamma^{\mathrm{D}})]$.

Region 3: High investor sentiment, $V_t^{\mathrm{D}} > V_t^{\mathrm{S}} + (Q/\gamma^{\mathrm{D}})$. In this region, only dumb investors participate in the market at time 2, and smart investors sit out. Prices are given by $P_t = V_t^{\mathrm{D}} - Q/\gamma^{\mathrm{D}}$.

The next step is to be more explicit about insider trading behavior at time 2, and the associated updating process that determines η^{E}. To do so, we follow Kyle (1985). The insider who observes η at time 2 is assumed to be rational and risk-neutral, and to trade by means of a market order. Unlike the outside investors, we allow the

[12] See, e.g., Chen et al. (2002), D'Avolio (2002), Geczy et al. (2002), Hong and Stein (2002), Lamont and Jones (2002) and Diether et al. (2002). Notable earlier papers include Harrison and Kreps (1978), Jarrow (1980), and Diamond and Verrecchia (1987).

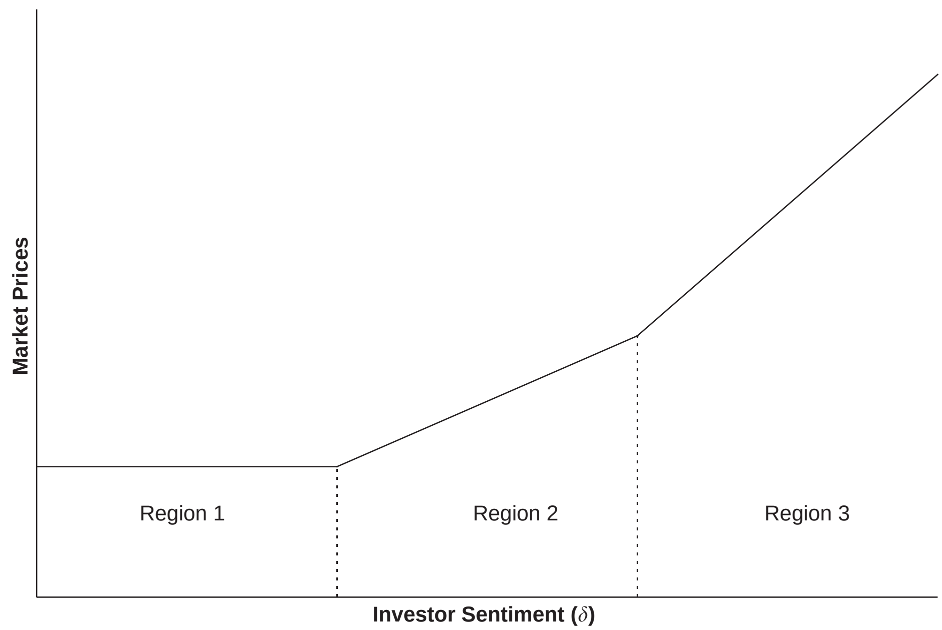

Fig. 1. Market pricing and investor sentiment. The relationship between market prices and the sentiment of dumb investors. There are three regions. In the first region, only smart investors participate. In the second region, both types of investors participate. In the third region, only dumb investors participate.

insider to go both long and short. His market order is absorbed by the pool of outside investors, who play a role analogous to that of Kyle's market-makers in this set-up. More precisely, since the insider's market order is of infinitesimal size, it only affects prices insofar as the information it contains alters outside investors' time-2 valuations, V_2^S and V_2^D. In addition to the insider, there are also some non-strategic liquidity traders active at time 2, who place exogenously given market orders in total amount z. The variance of z is also taken to be infinitesimally small, and for notational economy, we assume that it is the same as the variance of η.

While the insider observes η, he—unlike either type of outside investor—makes no attempt to estimate F. Nor does he have any direct knowledge of δ. To keep things especially simple, we assume that, prior to observing η, the insider's best estimate of the terminal dividend is simply the time-1 price P_1. In other words, the insider may have a tip about an upcoming earnings announcement, but he does not know anything else about fundamental value, and so just relies on P_1 as a summary statistic for the information about F that he does not have.[13]

[13] P_1 will in fact be the rational estimate of the terminal dividend for an agent who does not know F or δ if we choose the appropriate ex ante distribution for δ. For example, suppose that δ is symmetrically distributed, and takes on one of two values, each with probability one-half: either $\delta = Q/\gamma^S + Q/\gamma^D$; or $\delta = -Q/\gamma^S - Q/\gamma^D$. From Proposition 1, it follows that either $P_1 = F + Q/\gamma^S$ (Region 3); or $P_1 = F - Q/\gamma^S$ (Region 1). So P_1 is an unbiased estimator of F.

Let the size of the insider's market order be given by m. He seeks to maximize $\mathrm{E}\{m(F+\eta+\varepsilon-P_2)\}$, which, given that his estimate of F is P_1, can be written as

$$\max \mathrm{E}\{m(\eta-\Delta P_2)\}, \tag{3}$$

where $\Delta P_2=(P_2-P_1)$. Intuitively, the insider trades off exploiting his private information η against the adverse price impact of trade ΔP_2. The total order flow at time 2, which we denote by f, is given by $f=m+z$—i.e., the total order flow is the sum of that coming from the insider and the liquidity traders.

Given our assumptions, ΔP_2 can be written as

$$\Delta P_2 = w\eta^{\mathrm{E}}, \tag{4}$$

where w is an indicator variable that takes on the following values: $w=1$ in Region 1; $w=\theta+(1-\theta)\gamma^{\mathrm{S}}/(\gamma^{\mathrm{S}}+\gamma^{\mathrm{D}})$ in Region 2; and $w=\theta$ in Region 3. That is, the extent to which ΔP_2 reflects a full rational-expectations reaction to the new information available at time 2 depends on which region we are in, and hence on the sentiment parameter δ. As δ increases, and we move from Regions 1 to 3, the weight of the dumb traders in the pricing function increases, and ΔP_2 is progressively less influenced by η^{E}.

The rational-expectations revision η^{E} can in turn be pinned down from a regression of η on the order flow f

$$\eta^{\mathrm{E}} = \beta f, \tag{5}$$

where $\beta=\mathrm{cov}(\eta,f)/\mathrm{var}(f)$. So we can alternatively write ΔP_2 as

$$\Delta P_2 = w\beta f \equiv \lambda f, \tag{6}$$

where λ is the familiar Kyle (1985) depth parameter that measures the price impact of order flow. Note that here $\lambda=w\beta$, which contrasts with the standard fully rational version of the Kyle model, where the equivalent statement is simply that $\lambda=\beta$.

The insider seeks to maximize his objective function as stated in Eq. (3), taking the price-impact parameter λ as given. His first-order condition is therefore

$$m=\frac{\eta}{2\lambda}. \tag{7}$$

Given this expression for m, and the assumption that η and z have equal variance, we can compute the regression parameter β as

$$\beta=\frac{2\lambda}{1+4\lambda^2}. \tag{8}$$

Recalling that $\lambda=w\beta$, we have the following equilibrium condition for λ:

$$\lambda=\frac{2w\lambda}{1+4\lambda^2}. \tag{9}$$

Solving, we have that the equilibrium λ, which we denote as λ^*, is given by

$$\lambda^*=\sqrt{\frac{2w-1}{4}}. \tag{10}$$

Since w is decreasing as we move from Region 1 to Region 3, it follows from (10) that the price impact of a trade is also decreasing. Moreover, this decreased price impact leads to increased trading volume—faced with a more liquid market, the insider trades more aggressively.[14] To see this, note that a natural measure of expected time-2 trading volume, which we denote by T, is simply the variance of the order flow f:

$$T = \mathrm{var}(f) = \mathrm{var}(m) + \mathrm{var}(z) = \mathrm{var}(\eta)\frac{1 + 4\lambda^2}{4\lambda^2}. \tag{11}$$

We have thus established the following.

Proposition 2. *Liquidity increases with investor sentiment. As we move from Region* 1 *to Region* 2 *to Region* 3, *the market becomes more liquid at time* 2, *in the sense that the price impact of a trade decreases. Correspondingly*, *trading volume at time* 2 *also increases*.

The intuition for the proposition, which is illustrated in Fig. 2, is very simple. As can be seen from Eq. (10), the price impact of a trade is increasing in w, which is nothing more than a measure of the weight of the smart traders in the pricing function. In Region 1, where the smart traders dominate the market, w is high (it equals one) and hence liquidity and trading volume are low. At the other extreme, in Region 3, where the dumb traders dominate the market and smart traders are sitting on the sidelines, w is low (it equals θ) and hence liquidity and trading volume are high.

It is also worth pointing out that the results in Proposition 2 would be strengthened if we were to follow Admati and Pfleiderer (1988) and make the variance of liquidity trading endogenous, as opposed to keeping it a fixed exogenous constant. With optimizing liquidity traders, any decrease in λ brought about by a decline in w will tend to feed on itself—the lower price impact will induce liquidity traders to place more aggressive orders, thereby further reducing the information content of the order flow and causing a second-round multiplier decrease in λ.

The model's implications for the link between liquidity and expected returns now follow immediately. As dumb investors become more optimistic—as δ and hence V_t^{D} rise relative to V_t^{S}—not only do liquidity and trading volume increase, but expected returns fall. More precisely, we have that:

Proposition 3. *Expected returns are decreasing in liquidity. Define the expected return from time t until the terminal date*, R_t^{E}, *as* $R_t^{\mathrm{E}} = V_t^{\mathrm{S}} - P_t$. *For* $t = \{1, 2\}$, *the average value of* R_t^{E} *is decreasing in dumb-investor sentiment* δ *and hence decreasing as we move from Region* 1 *to Region* 2 *to Region* 3. *Thus*, *increases in market liquidity and trading*

[14] Note that in this version of the model, the dumb investors do not contribute directly to increased trading volume at time 2—they only play an indirect role, by making the market more liquid, and hence inducing the insider to trade more. Below, we discuss an extension of the model in which dumb investors have a direct effect on volume.

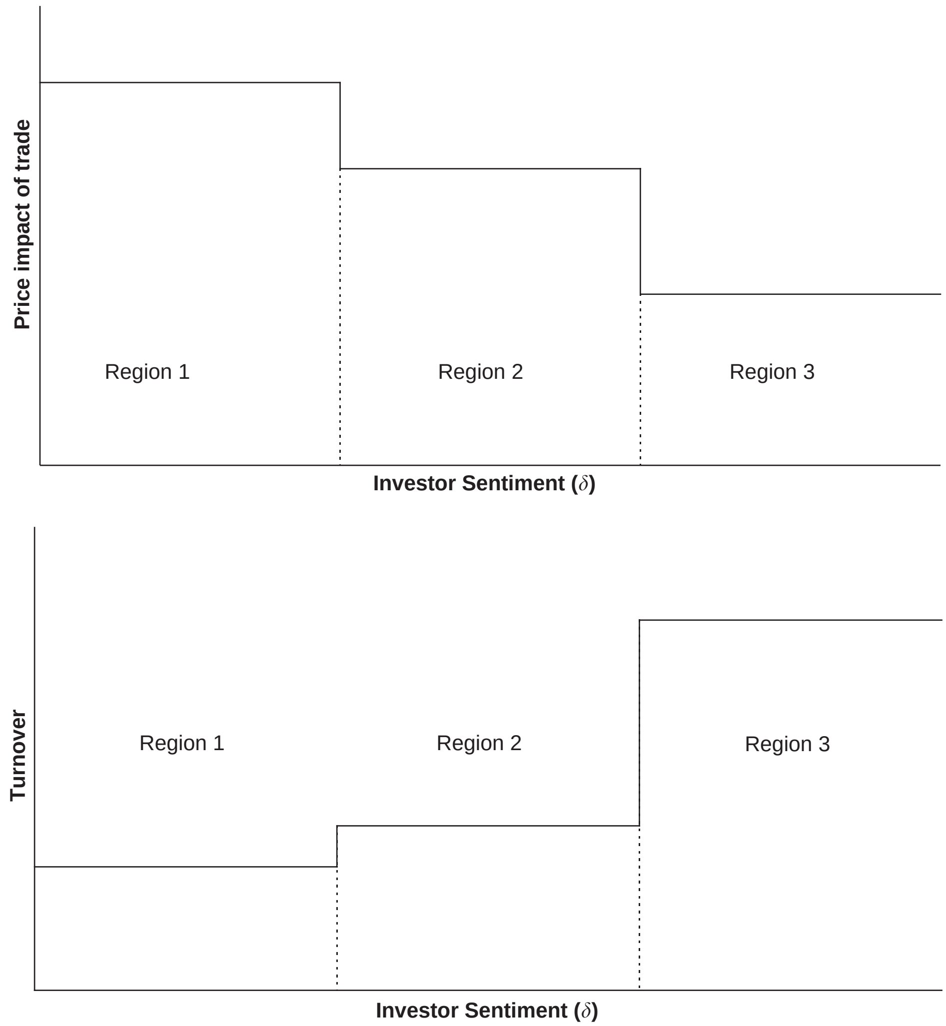

Fig. 2. Market liquidity and investor sentiment. The relationship between market liquidity and the sentiment of dumb investors. Market liquidity is measured with both the price impact of trade (first panel) and share turnover (second panel). There are three regions. In the first region, only smart investors participate. In the second region, both types of investors participate. In the third region, only dumb investors participate. The reaction of dumb investors to news (the parameter θ) is set to 0.6.

volume at time 2 *are associated with a reduction in subsequent expected returns—i.e., with a reduction in* R_2^{E}.

To see the key role played by the short-sales constraint, note that if this constraint is absent, it is as if we are always in Region 2, with both groups of investors active and the price given by $P_t = [\gamma^{\mathrm{S}}/(\gamma^{\mathrm{S}}+\gamma^{\mathrm{D}})]V_t^{\mathrm{S}} + [\gamma^{\mathrm{D}}/(\gamma^{\mathrm{S}}+\gamma^{\mathrm{D}})]V_t^{\mathrm{D}} - [Q/(\gamma^{\mathrm{S}}+\gamma^{\mathrm{D}})]$. This is now a standard noise-trader model, as in De Long et al. (1990). In such a case, it is still true that as dumb investors' sentiment goes up, expected returns fall, since dumb investors have a non-zero relative weight of $\gamma^{\mathrm{D}}/(\gamma^{\mathrm{S}}+\gamma^{\mathrm{D}})$ in the pricing formula. However, there is no longer any variation in liquidity or trading volume, since this relative weight is now a constant for all parameter values, and w is therefore also constant at $w=\theta+(1-\theta)\gamma^{\mathrm{S}}/(\gamma^{\mathrm{S}}+\gamma^{\mathrm{D}})$. Intuitively, even if dumb investors are much more optimistic than smart investors, and hence are doing all the buying, smart investors continue to exert the same marginal influence on price, by taking short positions.

By adding the short-sales constraint to our model, we create the following new effect: *as dumb investors become more optimistic*, *they drive the smart investors to the sidelines and hence gain a greater weight in the pricing function*. At the extreme, in Region 3, smart investors are completely out of the picture, and the dumb investors have a weight of unity in the pricing function. This greater weight, combined with the assumption that dumb investors underreact to the information contained in trades, leads to a more liquid market.

Another way to express the basic idea behind our model is to say that when the market is observed to be highly liquid, this suggests that it is currently being dominated by dumb investors—i.e., the inmates have taken over the asylum. And of course, in a world with short-sales constraints, the fact that the market is being dominated by dumb investors means that the sentiment of these dumb investors is positive, and hence that expected returns are relatively low.

3. A corporate-finance variation: equity issues and expected returns

It is easy to modify our model so that the notion of "liquidity" it captures is liquidity in the market for seasoned equity offerings (SEOs). We keep the basic structure and timing as before, and make a couple of changes. First, the insider who observes η at time 2 is no longer a trader, but instead is the manager of the firm, who is contemplating an (infinitesimally small) equity issue.[15] We assume that the manager's behavior can be summarized by a simple objective function. Specifically, he will issue equity if and only if

$$-\eta+\Delta P_2+K\geqslant 0, \tag{12}$$

where K is the net present value of the investment that can be financed with the equity issue. Thus the manager prefers to issue equity when his inside information η is unfavorable, when the adverse price impact of an issue ΔP_2 is small, and when the NPV of investment K is large.

This rendition of the manager's objectives is similar to that of Myers and Majluf (1984), with one crucial difference. In our formulation, *the manager does not attempt to make a comprehensive judgement of the firm's fundamental value—i.e.*, *he does not have an estimate of either F or δ*. Thus prior to observing η, the manager, like the

[15] Again, the reason for making the equity issue infinitesimally small is so that it does not affect the overall supply of shares outstanding, and hence does not change the equilibrium risk premium.

insider in the previous section, takes the price at time 1 to be a summary statistic for the expected terminal dividend.[16]

We use this formulation not because we believe it is necessarily the most realistic one. Perhaps managers do in fact have some comparative advantage in judging whether their firms are over- or undervalued relative to long-run fundamentals. But our goal is to show that even if they are not such astute market timers, and behave in a more simple rule-of-thumb fashion that ignores the relationship of the time-1 price to the fundamental F, their financing decisions can still forecast subsequent returns. So long as the rule of thumb contains an element of "issue equity when the price impact is small," financing decisions will be a mirror of market liquidity.[17] And in our framework, market liquidity forecasts returns for reasons that have nothing to do with managers being well informed about fundamentals.

The only other modification we make to the model of the previous section is to assume that η is now uniformly, rather than normally distributed.[18] As will become clear, this just simplifies the analysis. The support of η is on $[-x, x]$, and to avoid degenerate solutions where all types always issue equity regardless of their draw of η, we require that $K < x$.

The inequality in (12) can be re-written as

$$-\eta + w\eta^{\mathrm{E}} + K \geqslant 0, \tag{13}$$

where w is defined as before, and where η^{E} is now the rational expectation of η conditional on there being an equity issue.

Equilibrium in this version of the model consists of a threshold value of η, denoted by η^*, such that only a manager observing $\eta \leqslant \eta^*$ chooses to issue equity. If managers behave this way, then conditional on observing an equity issue, the rational inference is that

$$\eta^{\mathrm{E}} = \frac{\eta^* - x}{2}. \tag{14}$$

Plugging (14) into (13) and setting the inequality to zero, we can solve for η^*

$$\eta^* = \frac{2K - wx}{2 - w}. \tag{15}$$

Note that η^* is decreasing in w, since $\mathrm{d}\eta^*/\mathrm{d}w = (2K - 2x)/(2 - w)^2$, and $K < x$. Since η^* is effectively a measure of the intensity of equity issuance—the probability of an equity issue is given by $(\eta^* + x)/2x$—it follows that equity issuance increases as we move from Regions 1 to 3 and w falls. It is also easy to verify that the new issues

[16] As noted earlier, by picking the right ex-ante distribution for δ, we can make the time-1 price the best estimate of the terminal dividend for an agent who does not observe either F or δ.

[17] Such a rule of thumb may be implicitly encouraged by investment bankers. Butler et al. (2003) document that investment-banking fees in SEOs are lower for liquid firms. Presumably, this is because liquidity in the aftermarket facilitates the task of underwriting.

[18] Given that η has infinitesimal variance relative to ε, it is still the case that the terminal dividend is (approximately) normally distributed. So CARA utility still generates the sort of simple linear demand schedules and pricing relationships that we have been using.

market becomes more liquid, in the sense that the equilibrium price impact ΔP_2 becomes smaller, as we move from Regions 1 to 3. Thus we have established:

Proposition 4. *Expected returns are lower in hot issues markets. Expected returns from time* 2 *until the terminal date,* R_2^E, *are lower following hot issues markets*, *where a "hot" market is defined as one in which either*: (i) *more equity issues are observed at time* 2; *or* (ii) *the price impact of an equity issue at time* 2 *is smaller*.

Fig. 3 provides an illustration. Again, the striking feature to be emphasized is that financing decisions can forecast long-run returns even though managers themselves have no view about F, the dominant component of long-run fundamental value.

4. Discussion

We believe that the most attractive aspect of our theory is that it provides a unified framework for thinking about two quite distinct—and at first glance, unrelated—branches of empirical research: the body of work in market microstructure which seeks to relate measures of liquidity and trading volume to expected returns, and the corporate-finance literature on equity offerings and subsequent stock returns. Indeed, the theory can shed light on several more narrowly-defined topics within these two broad areas: (i) time-variation in firm-level liquidity and stock returns; (ii) the behavior of Internet stocks during that sector's boom; (iii) commonality in liquidity across firms and its link to aggregate stock returns; (iv) the firm-level new issues puzzle; and (v) economy-wide hot issue markets.

4.1. Firm-level liquidity and stock returns

The existing empirical evidence suggests that firm-level stock returns are increasing in the bid–ask spread, increasing in the price impact of trade, and decreasing in trading volume. Amihud and Mendelson (AM) (1986) sort firms into portfolios according to their bid–ask spread, once a year from 1961 to 1980. The beta-adjusted returns of the high-spread portfolio exceed those of the low-spread portfolio by 0.7 percent per month (AM Table 2). Similarly, Brennan and Subrahmanyam (BS) (1996) sort firms into portfolios in 1984 and 1988 according to an estimate of the Kyle (1985) price-impact parameter λ. The three-year average monthly returns for the high-λ firms (in the periods that follow) are higher than the returns on the low-λ firms by 0.6 to 1.4 percent per month (size-adjusted differences in BS Table 1). Brennan et al. (1998) find a negative relationship between lagged dollar volume and stock returns.[19] In each case, the economic significance is large, perhaps too large to

[19] By contrast, Gervais et al. (2001) find that firms that experience unusually high trading volume have *higher* subsequent returns. This "high volume premium" focuses on liquidity shocks rather than the level of liquidity and short-run rather than long-run returns, and so is not inconsistent with the basic relationship between high levels of turnover and low subsequent returns documented in Brennan et al. (1998).

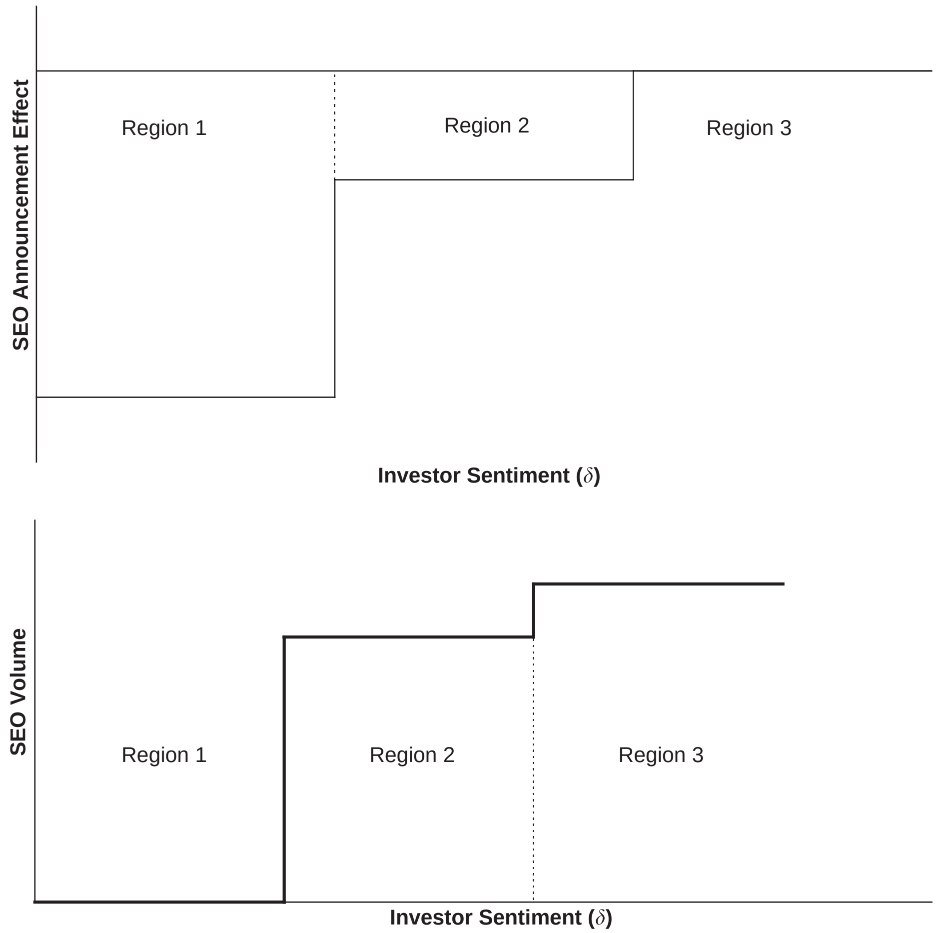

Fig. 3. New equity issues and investor sentiment. The relationship between equity issues and the sentiment of dumb investors. We measure both the price impact of an equity issue (first panel) and the volume of equity issues (second panel). There are three regions. In the first region, only smart investors participate. In the second region, both types of investors participate. In the third region, only dumb investors participate. The reaction of dumb investors to news (the parameter θ) is set to 0. We set the investment opportunity (K) equal to 0 and the support of the uniform distribution (x) equal to 1.

square with reasonable levels of turnover and a transaction-costs view of the liquidity premium.

It is important to note that each of these results has two components: the excess returns come from both within-firm time-series variation in liquidity as well as between-firm variation. Our theory can at most take partial credit for the former, while the latter is probably better explained with the traditional transaction-costs view. We clearly have little to say about why a firm whose stock is chronically illiquid has higher returns, year in and year out, than one whose stock is more liquid.

4.2. Behavior of Internet stocks during the boom

Many of the effects in our model are vividly illustrated by the behavior of Internet stocks during the boom period from January 1998 to February 2000. Ofek and Richardson (2003) document that the extraordinarily high valuations in this sector at this time were accompanied by very low bid–ask spreads and unusually high trading volume. For example, Ofek and Richardson report a median bid–ask spread for Internet firms of 0.5 percent over this period, compared with a statistically different 0.8 percent for non-Internet firms. Similarly, turnover for Internet firms was three times higher.

The low spreads in particular are hard to rationalize in the context of standard models of liquidity. From an inventory-management perspective, the high volatility of Internet stocks (with twice the variance of daily returns of non-Internet stocks) would seem to imply greater inventory risk for market makers, and hence wider spreads. And from an adverse-selection perspective, it is hard to imagine exogenous reasons why there should be less private information available to market participants about the prospects of Internet firms, especially given a flow rate of news in general (as proxied for by volatility) substantially higher than that in other industries.

In the context of our model, the explanation for the narrow spreads would begin with the premise that the Internet sector was greatly overvalued during this period. With short-sales constraints, this is equivalent to saying that the market for Internet stocks was dominated by irrational investors. These irrational investors were not apt to revise their valuations on the basis of subtle signals such as those embodied in order flow. In a revealing example of just this kind of underreaction, Schultz and Zaman (2001) and Meulbroek (2000) document that insider sales in Internet companies—unlike similar transactions in "old economy" firms—were not accompanied by negative stock-price impacts.[20] Faced with such stickiness in valuations, it seems plausible that market makers could safely quote narrow spreads, confident that any resulting inventory imbalances they had to offload would be absorbed with minimal price concessions.

Our model can also explain the high turnover of Internet stocks with a corollary logic: given the tighter spreads, it became cheaper for investors to trade, and so they did more of it. However, this part of the story may be a bit hard to take literally, since it asserts that *all* of the increase in turnover came from a reduction in the *costs* of trading, as opposed to an outward shift in the *demand* to trade. Casual empiricism suggests that increased trading demand must also have played an important role during the Internet boom.

It is straightforward to extend the model to capture this trading-demand effect. The key to doing so is to assume that dumb investors are more prone to churning

[20] Of course, there are several possible interpretations of this finding. The first is that there was little asymmetric information at this time in this industry. The second is that Internet insiders were particularly undiversified and thus had other, non-information-related reasons to sell stock. And the third is that private information was simply not incorporated into prices in the process of trading, consistent with our Proposition 2.

their positions than are smart investors. To be more specific, keep aggregate dumb-investor demand exactly as before at both time 1 and time 2, but introduce some small divergences in the relative valuations of dumb investors at time 2 only. In the context of our overconfidence framework, this amounts to assuming that dumb investors now get additional private signals at time 2, some of which are positive, and some of which are negative.[21] This implies that if dumb investors are long to begin with at time 1—i.e., we are in Region 2 or 3 of the model—they trade amongst themselves at time 2, even absent any orders from the insider. As before, the crucial element is the short-sales constraint, which implies that those investors who contribute the most to market liquidity (either by underreacting to the information in order flow, or by having a greater desire to trade among themselves) are only present when prices are relatively high.[22]

4.3. Commonality in liquidity and aggregate stock returns

Recent research suggests that common market-wide factors drive firm-level liquidity. Chordia, Roll and Subrahmanyam (CRSa) (2000) show that quoted spreads, depth, and effective spreads for NYSE firms move with the time-series of the across-firm averages (CRSa Table 3). In addition, quoted spreads are strongly negatively related to market and industry turnover (CRSa Table 8). Huberman and Halka (2001) and Hasbrouck and Seppi (2001) come to similar conclusions with different underlying data. Using a different approach, Lo and Wang (2000) find that a single factor explains as much as 80 percent of the variation in turnover in a cross-section of stock portfolios.

Chordia et al. (CRSb) (2001) identify some of the common market-wide factors in liquidity, finding that changes in spreads, depths, and turnover respond to short-term interest rates, the term spread, and past market returns and volatility (CRSb Table 5). None of these models explains more than about a third of the variation in liquidity, however.

Consistent with our model as well as with the traditional view of liquidity, Amihud (2002) and Jones (2002) find that market turnover, the ratio of the absolute market return to turnover (a rough notion of price impact), and the average bid–ask spread are good predictors of future returns. The distinguishing empirical prediction of our model is the extent of this predictability, which we evaluate below. While rational

[21] This extension of the model resembles the mechanism in Scheinkman and Xiong (2003). However, their model has *only* overconfident types, and no rational investors, which makes it more awkward for them to rationalize the Internet-boom feature of high prices accompanied by extraordinary volume—to do so, they effectively need to assume that the variance of private signals was abnormally high during the boom. In our setting, we can be more parsimonious, and keep the variance of all shocks constant.

[22] Once we add the ingredient that dumb investors have a greater desire to trade among themselves than smart investors, we no longer need to assume that they underreact to the information in order flow, if the only goal is to generate a link from turnover to prices. However, we still need the underreaction assumption for our results on equity issues. The price reaction to an equity issue does not depend directly on how much volume there is at time 2, since equity issues are not pooled with general order flow.

transaction-cost theories would seem to suggest relatively modest effects, investor-sentiment-driven movements in liquidity can in principle be associated with greater predictability.

4.4. The new issues puzzle

A long list of papers, including Stigler (1964), Ritter (1991), Loughran and Ritter (1995), and Speiss and Affleck-Graves (1995), document that issuing firms earn low returns relative to market benchmarks. There is some question as to whether the market overall is a good benchmark. Issuers tend to be small and have high ratios of market to book value, a combination of characteristics that Brav and Gompers (1997) connect with low returns among non-issuers. And Fama (1998) and Mitchell and Stafford (2000) raise additional questions of economic significance (value-weighted average returns are higher) and statistical significance (issuing activity is clustered in time and so issuers' returns are not independent). However, nobody disputes that the typical returns are low. In Brav and Gompers (1997), the average issuing firm performs worse than Treasury bills.

Low post-issue returns are consistent with both liquidity-motivated and market-timing theories of equity issuance. However, Eckbo, Masulis and Norli (EMN) (2000) offer micro evidence that suggests that liquidity is at least part of the story. They document that issuers, all else equal, tend to be more liquid than non-issuers: issuers' turnover in the period up to five years prior to an SEO is a third higher than that of size and book-to-market matched firms (EMN Table 11). Again, this seems to fit with the key ideas in our model: that financing decisions may be made on the basis of the market's current willingness to absorb new issues; and that this form of liquidity, like spreads, depth, and trading volume, carries information about the extent to which irrational traders are influential in the market, and hence about expected returns.

4.5. Hot issue markets and aggregate stock returns

Like liquidity, equity issuance varies dramatically over time. Bayless and Chaplinsky (1996) find that "hot issue markets"—i.e., periods of heavy SEO activity—coincide with a reduced average price impact of issue. By itself, this finding could be perfectly well explained in a fully rational model with time-varying adverse selection; indeed just such an approach is taken by Choe et al. (1993). But a rational adverse-selection-based model cannot explain the important other side of the coin, namely that, as shown by Baker and Wurgler (2000), hot issue markets also portend low future market-wide returns. Conversely, a simple story whereby smart managers exploit dumb investors may rationalize the Baker–Wurgler finding, but such a story has nothing to say about time-variation in announcement event impacts. Our model offers a unified interpretation for these aggregate facts about hot issue markets, and does so without relying on the assumption that managers are better judges of long-run value than the average investor.

5. Evidence on aggregate turnover, equity issurance and stock returns

We now turn to a brief examination of the aggregate data on share turnover, equity issues and stock returns. As discussed in the Introduction, our aim is not to provide a sharp test of the model, but rather to document some broad-brush facts that are suggestively consistent with it. The one potential wedge between the traditional view of liquidity and our model is the economic significance of liquidity as a predictor of future returns. Examining this predictive power, and documenting the surprisingly strong correlation between turnover and new equity issues, are the two main goals of this section.

5.1. Data

We generate an annual series on turnover by taking the ratio of reported NYSE share volume to average shares listed; both of these components come from the NYSE *Fact Book*, and are available since 1900.[23] Our measure of equity issuance is the ratio of common and preferred issues in a given year to the sum of these two items plus public and private debt issues; these items are from the *Federal Reserve Bulletin* and are available since 1927. Baker and Wurgler (2000) provide a more detailed description of how the equity share is constructed. Our stock market returns are those on the CRSP value-weighted and equal-weighted portfolios, converted to real terms with the consumer price index from Ibbotson (2001). We control for general valuation levels using the corresponding CRSP dividend yield. The binding constraint is the availability of the equity share data, so our full sample period runs from 1927 to 1998.

5.2. Turnover and the equity share in new issues

Fig. 4 plots the level of turnover and the equity share over this period. The two series generally appear to move together closely, though there is a pronounced break in the turnover series that roughly coincides with the so-called "Big Bang" deregulation of brokerage commissions in 1975. In May of that year, a Securities and Exchange Commission (SEC) ruling prevented securities exchanges from fixing brokerage commission rates.[24] Afterward, competition intensified, prices fell, and turnover increased. In another apparent structural break, the equity share declines dramatically (and separately from turnover) in the mid-1980s. Baker and Wurgler (2000) attribute the six-fold real increase in debt issues between 1982 and 1986 partly to lower interest rates and a growing market for junk bonds. In addition, Baker et al.

[23] We unfortunately do not have access to other measures of aggregate market liquidity over this long a sample period. In Jones' (2002) proprietary data set, market-wide turnover and bid–ask spreads seem to capture a good deal of common information—the correlation of annual changes is—0.40 over the period 1900–1998.

[24] See Ofer and Melnik (1978) for more detail on the SEC ruling.

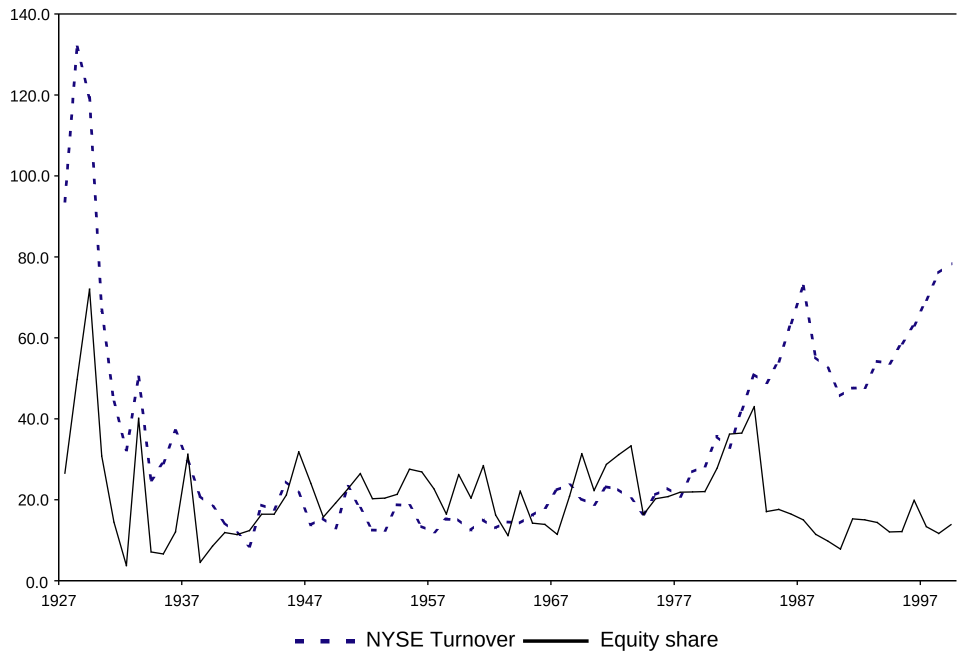

Fig. 4. Market liquidity and equity issues. Market liquidity is the ratio of reported share volume to average shares listed times 100 from the NYSE *Fact Book*. Equity issues are scaled by total equity and debt issues, multiplied by 100, and plotted on the same axis as market liquidity. Equity and debt issue volumes are from the *Federal Reserve Bulletin*. Equity includes both common and preferred equity issues. Debt includes both public and private debt issues. NYSE market capitalization is from CRSP.

(2002) document a declining maturity of new debt issues prior to this period, leading to more rapid refinancing.

In light of these sharp trends in the latter part of the sample period, we often work with stochastically detrended versions of our variables, where the stochastic detrending is done very simply, by subtracting the mean of the previous five years' realizations from the current value. Gallant et al. (1992) and Andersen (1996) advocate this sort of detrending in share turnover data. Again, because of the constraint on the equity share data, this further reduces our sample (by five years) to the period from 1932 to 1998. Lo and Wang (2000, 2001) prefer to use levels of turnover and instead divide the sample into subperiods. So we also do some limited experimentation with a shorter, pre-Big-Bang sample period of 1927 to 1974.

Table 1 presents summary statistics for our raw data, presented in levels in the first four columns and in detrended levels in the second four columns. Panel A shows the pre-1975 sample, and Panel B shows the full sample. In the full sample, turnover averages about 24 percent per year, though the average in the last ten years has been much higher, at 59 percent per year. The standard deviation of the level of turnover is 25 percent. Detrended turnover is less variable, with a standard deviation of 12 percent, in part because detrending leaves us with a shorter sample period that begins

Table 1
Correlations among market liquidity, the equity share in new issues and valuation ratios

	Levels				Detrended levels			
	Turnover	S	VW D/P	EW D/P	Turnover	S	VW D/P	EW D/P
Panel A:	1927–1974				1932–1974			
N	48	48	48	48	43	43	43	43
Mean	26.83	21.69	4.56	4.01	−4.22	−0.70	−0.10	−0.10
SD	25.61	11.91	1.32	1.50	12.29	9.82	1.07	1.18
Correlation								
Turnover	1.00				1.00			
$S = e/(e+d)$	0.64	1.00			0.67	1.00		
$VW\ D/P$	−0.06	−0.27	1.00		−0.34	−0.31	1.00	
$EW\ D/P$	−0.18	−0.12	0.86	1.00	−0.07	−0.05	0.87	1.00
Panel B:	1927–1998				1932–1998			
N	72	72	72	72	67	67	67	67
Mean	33.80	20.84	4.25	3.40	−0.55	−0.99	−0.14	−0.12
SD	24.83	11.00	1.32	1.55	12.28	9.24	0.94	0.97
Correlation								
Turnover	1.00				1.00			
$S = e/(e+d)$	0.36	1.00			0.44	1.00		
$VW\ D/P$	−0.31	−0.03	1.00		−0.33	−0.21	1.00	
$EW\ D/P$	−0.43	0.04	0.84	1.00	−0.14	−0.04	0.84	1.00

Univariate relationships among market liquidity, the equity share in new issues, and the dividend yield. Market liquidity is the ratio of reported share volume to average shares listed from the NYSE *Fact Book*. Equity and debt issue volumes are from the *Federal Reserve Bulletin*. Equity includes both common and preferred equity issues. Debt includes both public and private debt issues. The dividend yields (D/P) are calculated separately on the CRSP value-weighted (VW) and equal-weighted (EW) portfolios.

in 1932 and that thus excludes the volatile years around the crash of 1929. The mean share of equity to total new equity and debt issues is about 20 percent. All of the detrended means are slightly below zero, indicating a low frequency downward trend in turnover, equity issues, and dividend yields from 1927 to 1998.

Turnover and the equity share are highly correlated. The lowest correlation coefficient is 0.36, for the full sample period in levels. In the pre-1975 sample, the correlation is as high as 0.67. Both of these variables are negatively correlated with dividend yields. In other words, when valuations are high relative to dividends, so too are liquidity and equity issues. These correlations are generally slightly stronger in the detrended data and with the value-weighted dividend yield.

We examine the relationship between equity issues and turnover somewhat more formally in Table 2, regressing the equity share on contemporaneous turnover and the dividend yield. We also include three years of past returns as additional controls, given the well-known tendency for turnover to be related to past returns (Shefrin and Statman, 1985; Lakonishok and Smidt, 1986; Odean, 1998b). Our regression

Table 2
Market liquidity and the equity share in new issues

	Levels				Detrended levels			
	Coef	(t-stat)	Coef	(t-stat)	Coef	(t-stat)	Coef	(t-stat)
Panel A:	1927–1974				1932–1974			
Turnover	7.68	[4.89]	9.59	[5.71]	6.61	[8.63]	6.73	[5.14]
$VW\ D/P$			−1.85	[−1.26]			−0.31	[−0.16]
R_{et-1}			2.46	[1.58]			1.37	[0.68]
R_{et-2}			0.79	[0.64]			−0.69	[−0.41]
R_{et-3}			−0.36	[−0.26]			−1.27	[−1.39]
N		48		46		43		43
R^2		0.42		0.50		0.45		0.50
Panel B:	1927–1998				1932–1998			
Turnover	4.00	[1.50]	3.30	[0.93]	4.10	[2.71]	3.84	[2.24]
$VW\ D/P$			1.55	[0.82]			−0.05	[−0.04]
R_{et-1}			2.40	[1.78]			1.45	[1.01]
R_{et-2}			0.60	[0.45]			−0.30	[−0.26]
R_{et-3}			−0.17	[−0.17]			−0.03	[−0.03]
N		72		70		67		67
R^2		0.13		0.12		0.20		0.22

OLS regressions of the equity share in new issues on market liquidity and valuation ratios:

$$S_t = a + b\,\text{Turnover}_t + c\frac{D}{P_t} + d_1 R_{t-1} + d_2 R_{t-2} + d_3 R_{t-3} + u_t,$$

where S denotes the equity share in new issues. Equity and debt issue volumes are from the *Federal Reserve Bulletin*. Equity includes both common and preferred equity issues. Debt includes both public and private debt issues. Market liquidity (*Turnover*) is the ratio of reported share volume to average shares listed from the NYSE *Fact Book*. The dividend yield (D/P) is calculated on the CRSP value-weighted (VW) portfolio. The independent variables are standardized within each subperiod to have unit variance. The first four columns show regressions in levels. The second four columns show stochastically detrended results, where the prior five-year mean is subtracted from the current level of the dependent and independent variables. T-statistics are in braces and use two-lag Newey–West standard errors.

specification is thus

$$S_t = a + b\,\text{Turnover}_t + c\frac{D}{P_t} + d_1 R_{t-1} + d_2 R_{t-2} + d_3 R_{t-3} + u_t, \tag{16}$$

where S is the equity share, D/P is the CRSP value-weighted dividend yield, and R is the return on the CRSP value-weighted market portfolio.

The strong univariate correlation between turnover and the equity share from Table 1 generally holds up well in this multivariate setting. The only exceptions occur, not too surprisingly, when we use the full sample period and fail to detrend the data. When we either restrict attention to the pre-1975 sample period, or detrend the data (or both), the relationship is quite economically significant, even controlling for the dividend yield and past returns. For example, depending on the exact specification, a one-standard-deviation change in turnover leads to an increase in the equity share of between 4 and 10 percent (so that the equity share rises from, say, its sample mean of 20 percent to between 24 and 30 percent).

5.3. Turnover and the equity share as predictors of future returns

Next, in Table 3, we use turnover and the equity share to predict one-year-ahead real value- and equal-weighted returns, while controlling for the known influence of the dividend yield. Our regressions are all variants on the following general specification:

$$R_t = a + bS_{t-1} + c\,\mathrm{Turnover}_{t-1} + d\frac{D}{P_{t-1}} + u_t, \tag{17}$$

though in some cases we look at univariate or bivariate versions of the specification, effectively setting subsets of the coefficients b, c and d to zero. In contrast to the previous two tables, Table 3 restricts attention to the full sample period, which means that returns are measured over the interval 1933–1999. In unreported tests, we

Table 3
Multivariate regressions for predicting one-year-ahead market returns, 1933–1999

	Value-Weighted				Equal-Weighted			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
S	−7.21		−6.24	−6.11	−13.37		−8.96	−9.05
	−7.25		−6.45	−6.37	−13.37		−9.09	−9.26
	[0.01]		[0.02]	[0.02]	[0.00]		[0.02]	[0.02]
Turnover		−5.01	−2.21	−1.73		−14.07	−10.05	−9.49
		−4.25	−1.33	−0.99		−12.84	−8.82	−8.42
		[0.18]	[0.53]	[0.64]		[0.02]	[0.07]	[0.10]
D/P				1.63				3.69
				0.90				2.60
				[0.54]				[0.32]
N	67	67	67	67	67	67	67	67
R^2	0.15	0.07	0.16	0.16	0.20	0.21	0.29	0.30

Regressions of annual real equity market returns on the equity share in new issues, market liquidity, and valuation ratios:

$$R_t = a + bS_{t-1} + c\,\mathrm{Turnover}_{t-1} + d\frac{D}{P_{t-1}} + u_t,$$

where R denotes real percentage returns on the CRSP value-weighted (VW) or equal-weighted (EW) portfolio. Equity and debt issue volumes are from the *Federal Reserve Bulletin*. Equity includes both common and preferred equity issues. Debt includes both public and private debt issues. Market liquidity is the ratio of reported share volume to average shares listed from the NYSE *Fact Book*. The dividend yields (D/P) are calculated separately on the CRSP value-weighted (VW) and equal-weighted (EW) portfolios. The independent variables are standardized within each subperiod to have unit variance and stochastically detrended. We report OLS coefficients and small sample bootstrap, bias-adjusted coefficients below. Bootstrap p-values are in brackets, and represent a two-tailed test of the null hypothesis of no predictability.

find that the point estimates of the turnover coefficient c are generally higher in the pre-1975 subsample, but less precisely estimated.

Unlike Baker and Wurgler (2000) and Jones (2002), we use detrended levels of the equity share and turnover. For the equity share, this detrending makes little difference. However, for turnover, which is more persistent, detrending makes a considerable difference in forecasting power. The results shown here are stronger than Jones (2002) for this reason, and for two additional reasons: (i) we use a shorter sample period that excludes the very volatile pre-1927 turnover data; and, (ii) we present results with an equal-weighted index, thus giving more emphasis to the impact of turnover on the returns of small stocks.

A caveat here is that using a small sample and OLS regressions to forecast stock returns can lead to biased coefficients. Stambaugh (1999) shows that the magnitude of the bias depends on the persistence of the explanatory variables and the contemporaneous correlation between innovations in the explanatory variables and stock returns. For example, in a univariate model given by

$$R_t = a + bX_{t-1} + u_t, \tag{18}$$

$$X_t = c + dX_{t-1} + v_t. \tag{19}$$

Stambaugh shows that the bias is equal to

$$\mathrm{E}[\hat{b} - b] = \frac{\sigma_{uv}}{\sigma_v^2}\,\mathrm{E}[\hat{d} - d], \tag{20}$$

where the hats represent OLS estimates. The first term on the right-hand side of (20) is increasing in the contemporaneous correlation between changes in the predictor X and returns R, and the second term is increasing in absolute value in the degree of persistence in the predictor d.[25] For the equity share, the bias is small, because both d and σ_{uv}/σ_v^2 are small. For turnover and the dividend yield, we cannot dismiss the problem so easily. In both cases, d is large and u and v are highly correlated.

To deal with the problem, we use a bootstrap estimation technique. The approach closely resembles Vuolteenaho (2000), but it is also similar in spirit to Kothari and Shanken (1997), Stambaugh (1999), and Ang and Bekaert (2001). For each regression, we perform two sets of simulations, the first to generate a bias-adjusted point estimate, the second to generate a p-value that corresponds to the probability of observing the OLS point estimate under the null of no predictability. In the first set, we simulate (18) and (19) recursively starting with X_0, using the OLS coefficient estimates, and drawing with replacement from the empirical distribution of the errors u and v. We throw out the first 100 draws, drawing an additional N observations, where N is the size of the original sample.[26] With each simulated sample, we re-estimate (18). This gives us a set of coefficients b^*. Our bias-adjusted coefficient then subtracts the bootstrap bias estimate (which is the mean of b^* minus the OLS estimate of b) from the OLS b.

[25] Kendall (1954) shows that when d is large, the OLS estimate of d is biased downward.

[26] The effect of throwing out the first 100 draws is to draw from the unconditional distribution of X.

In the second set of simulations, we redo everything as before, except under the null hypothesis of no predictability—i.e., we impose the restriction that b is equal to zero. This gives us a second set of coefficients b^{**}. With these in hand, we can determine the probability of observing an estimate as large as the OLS b by chance, when the true b is equal to 0—this is where the p-values we report come from. In the multivariate regressions, we need a separate simulation for each predictor. In each case, the null hypothesis is no *marginal* predictive power for that variable.

Table 3 shows the results of our forecasting exercise. In a univariate regression, we find that a one-standard-deviation increase in detrended turnover leads to a reduction in year-ahead value-weighted returns of four percent and to a reduction in year-ahead equal-weighted returns of 13 percent.[27] The bootstrap standard errors are large, so the value-weighted results are not statistically significant. However, the equal-weighted estimates are significant at the two percent level.[28] The univariate results for the equity share imply similar economic magnitudes, but the standard errors are a good deal smaller, so these results are statistically significant in all cases.

When we look at the multiple regressions that include both turnover and the equity share simultaneously (along with the dividend yield), the coefficient on each drops noticeably, which is not surprising given their strong positive correlation with one another. However, for equal-weighted returns, both turnover and the equity share retain an economically meaningful independent effect: the incremental impact of a one-standard-deviation increase in either variable is to reduce year-ahead expected returns by roughly nine percent, and the two variables together produce a strikingly large OLS R^2 of 29 percent.[29] One interpretation is that both of these variables capture a component of "true" market liquidity, and by extension, a component of underlying investor sentiment.

To put the magnitudes in Table 3 into perspective, note that Jones (2002) finds that the standard deviation of commissions plus the bid–ask spread is 0.43 percent in the period from 1900 to 1998. According to a traditional theory of liquidity premia, this time-series variation in trading costs would have to explain the large time-series variation in expected returns that we document. Given that turnover is almost always less than 100 percent per year, it is hard to see why a rational representative investor would react to a *partially transitory* 0.43 percentage-point increase in trading costs by discounting stock prices to the point that they return an additional several percent over the next year alone.

[27] These bias-adjusted coefficients are noticeably lower than the OLS point estimates. Because turnover is persistent and its innovations are contemporaneously correlated with returns, this bias is anticipated. The equity share, by contrast, does not share these properties and therefore has little bias in its OLS coefficients.

[28] In unreported regressions, we find that the equal-weighted results are quite sensitive to sample period used. For example, the 13 percent figure rises to 17 percent when we focus on the pre-1975 period, but falls to 6 percent when we exclude the extraordinarily high equal-weighted return of 139 percent in 1933.

[29] The multivariate estimates for turnover are now only significant at the 10 percent level, however.

6. Conclusions

The basic idea of this paper is that, in a world with short-sales constraints, market liquidity can be a sentiment indicator. An unusually liquid market is one in which pricing is being dominated by irrational investors, who tend to underreact to the information embodied in either order flow or equity issues. Thus high liquidity is a sign that the sentiment of these irrational investors is positive, and that expected returns are therefore abnormally low.

The model we have used to formalize this idea is admittedly very simplistic. For example, it lacks any real dynamic element, and hence cannot speak to issues such as the horizon over which return predictability plays itself out. The model also requires—in addition to the short-sales constraints—a strong assumption, namely that the same investors who are subject to sentiment swings are also the most prone to underreact to certain kinds of subtle news. While one can appeal to a variety of a priori arguments and experimental evidence to motivate the plausibility of this assumption, we believe that our use of it is ultimately best defended on the grounds of the explanatory mileage that it yields.

In particular, the model is able to provide a unified explanation for a wide range of liquidity-related phenomena in stock markets. Many of the individual findings—from the return-forecasting power of measures of trading activity and trading costs, to the new issues puzzle and the existence of hot issue markets—have heretofore been rationalized separately, each with a story of its own. But as our preliminary empirical work suggests, these facts are intimately related to one another. So it is natural to want to be able to understand them within the context of a single conceptual framework. This paper has been a first attempt at developing such a framework; it would seem that there is room for much more to be done in this vein.

Ranging further afield, one might ask whether our liquidity-as-sentiment approach can also shed some light on the workings of other, more "real" asset markets, such as those for physical corporate assets or for houses. Many of these real markets are also characterized by a strong link between prices and measures of both trading volume and liquidity. This link has been studied by Shleifer and Vishny (1992), Stein (1995), and Pulvino (1998), all of whom assume rational investors and emphasize instead the roles of borrowing constraints and asset specificity. But perhaps investor sentiment also has some part to play in explaining the joint behavior of prices and liquidity in these other types of asset markets.[30] It would be interesting to develop this conjecture more completely, and to see whether it yields any novel empirical predictions.

References

Admati, A., Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. Review of Financial Studies 1, 3–40.

[30] At a minimum, these markets satisfy a necessary condition of our model, in that shorting is essentially impossible. Indeed, on this score, the real markets are a better fit to our assumptions than is the stock market.

Almazan, A., Brown, K., Carlson, M., Chapman, D., 2003. Why constrain your mutual fund manager. Journal of Financial Economics, forthcoming.

Amihud, Y., 2002. Illiquidity and stock returns. Journal of Financial Markets 5, 31–56.

Amihud, Y., Mendelson, H., 1980. Dealership market: market-making with inventory. Journal of Financial Economics 8, 31–53.

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid–ask spread. Journal of Financial Economics 17, 223–249.

Andersen, T., 1996. Return volatility and trading volume: an information flow interpretation. Journal of Finance 51, 169–204.

Ang, A., Bekaert, G., 2001. Stock return predictability: is it there? NBER working paper #8207.

Baker, M., Wurgler, J., 2000. The equity share in new issues and aggregate stock returns. Journal of Finance 55, 2219–2257.

Baker, M., Greenwood, R., Wurgler, J., 2002. The maturity of debt issues and predictable variation in bond returns. Journal of Financial Economics, forthcoming.

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343.

Bayless, M., Chaplinsky, S., 1996. Is there a window of opportunity for seasoned equity issuance? Journal of Finance 51, 253–278.

Bernard, V., Thomas, J., 1989. Post-earnings-announcement drift: delayed price response or risk premium? Journal of Accounting Research 27, 1–36.

Bernard, V., Thomas, J., 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. Journal of Accounting and Economics 13, 305–340.

Brav, A., Gompers, P., 1997. Myth or reality? The long-run underperformance of initial public offerings: evidence from venture capital and nonventure capital-backed companies. Journal of Finance 52, 1791–1822.

Brennan, M., Subrahmanyam, A., 1996. Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. Journal of Financial Economics 41, 441–464.

Brennan, M., Chordia, T., Subrahmanyam, A., 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. Journal of Financial Economics 49, 345–373.

Butler, A., Grullon, G., Weston, J., 2003. Does stock market liquidity matter? Evidence from seasoned equity offerings. Unpublished working paper, Rice University.

Chen, J., Hong, H., Stein, J., 2002. Breadth of ownership and stock returns. Journal of Financial Economics 66, 171–205.

Choe, H., Masulis, R., Nanda, V., 1993. Common stock offerings across the business cycle: theory and evidence. Journal of Empirical Finance 1, 3–31.

Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. Journal of Financial Economics 56, 3–28.

Chordia, T., Roll, R., Subrahmanyam, A., 2001. Market liquidity and trading activity. Journal of Finance 56, 501–530.

Copeland, T., Galai, D., 1983. Information effects on the bid–ask spread. Journal of Finance 38, 1457–1469.

D'Avolio, G., 2002. The market for borrowing stock. Journal of Financial Economics 66, 271–306.

De Long, J., Shleifer, A., Summers, L., Waldmann, R., 1990. Noise trader risk in financial markets. Journal of Political Economy 98, 703–738.

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

Diamond, D., Verrecchia, R., 1987. Constraints on short-selling and asset price adjustment to private information. Journal of Financial Economics 18, 277–311.

Diether, K., Malloy, C., Scherbina, A., 2002. Differences of opinion and the cross-section of stock returns. Journal of Finance 57, 2113–2141.

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

Eckbo, B., Masulis, R., Norli, O., 2000. Seasoned public offerings: resolution of the "new issues puzzle". Journal of Financial Economics 56, 251–292.

Edwards, W., 1968. Conservatism in human information processing. In: Kleinmutz, B. (Ed.), Formal Representation of Human Judgment, Wiley, New York.
Fama, E., 1998. Market efficiency, long-term returns, and behavioral finance. Journal of Financial Economics 49, 283–306.
Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81, 607–636.
Gallant, R., Rossi, P., Tauchen, G., 1992. Stock prices and volume. Review of Financial Studies 5, 199–242.
Garman, M., 1976. Market microstructure. Journal of Financial Economics 3, 257–275.
Geczy, C., Musto, D., Reed, A., 2002. Stocks are special too: an analysis of the equity lending market. Journal of Financial Economics 66, 241–269.
Gervais, S., Kaniel, R., Mingelgrin, D., 2001. The high-volume return premium. Journal of Finance 56, 877–919.
Glosten, L., Milgrom, P., 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.
Graham, J., Harvey, C., 2001. The theory and practice of corporate finance: evidence from the field. Journal of Financial Economics 60, 187–243.
Grossman, S., Miller, M., 1988. Liquidity and market structure. Journal of Finance 43, 617–637.
Harrison, J., Kreps, D., 1978. Speculative investor behavior in a stock market with heterogeneous expectations. Quarterly Journal of Economics 92, 323–336.
Hasbrouck, J., Seppi, D., 2001. Common factors in prices, order flows, and liquidity. Journal of Financial Economics 59, 383–411.
Hong, H., Stein, J., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. Journal of Finance 54, 2143–2184.
Hong, H., Stein, J., 2002. Differences of opinion, short-sales constraints and market crashes. Review of Financial Studies 16, 487–525.
Hong, H., Lim, T., Stein, J., 2000. Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. Journal of Finance 55, 265–295.
Huberman, G., Halka, D., 2001. Systematic liquidity. Journal of Financial Research 24, 161–178.
Ibbotson Associates, 2001. Stocks, Bonds, Bills, and Inflation. Ibbotson Associates, Chicago.
Jarrow, R., 1980. Heterogeneous expectations, restrictions on short sales, and equilibrium asset prices. Journal of Finance 35, 1105–1113.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.
Jones, C., 2002. A century of stock market liquidity and trading costs. Unpublished working paper, Columbia University.
Kendall, M., 1954. Note on bias in estimation of auto-correlation. Biometrika 41, 403–404.
Klibanoff, P., Lamont, O., Wizman, T., 1998. Investor reaction to salient news in closed-end country funds. Journal of Finance 53, 673–699.
Kothari, S., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: a time-series analysis. Journal of Financial Economics 44, 169–203.
Kyle, A., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.
Kyle, A., Wang, F., 1997. Speculation duopoly with agreement to disagree: can overconfidence survive the market test? Journal of Finance 52, 2073–2090.
Lakonishok, J., Smidt, S., 1986. Volume for winners and losers: taxation and other motives for stock trading. Journal of Finance 41, 951–974.
Lamont, O., Jones, C., 2002. Short sale constraints and stock returns. Journal of Financial Economics 66, 207–239.
Lo, A., Wang, J., 2000. Trading volume: definitions, data analysis, and implications of portfolio theory. Review of Financial Studies 13, 257–300.
Lo, A., Wang, J., 2001. Trading volume: implications of an intertemporal capital asset pricing model. Unpublished working paper, MIT.
Loughran, T., Ritter, J., 1995. The new issues puzzle. Journal of Finance 50, 23–51.

Meulbroek, L., 2000. Does risk matter? Corporate insider transactions in Internet-based firms. Unpublished working paper, Harvard University.

Mitchell, M., Stafford, E., 2000. Managerial decisions and long-term stock price performance. Journal of Business 73, 287–329.

Myers, S., Majluf, N., 1984. Corporate financing and investment decisions when firms have information that investors do not have. Journal of Financial Economics 13, 187–221.

Odean, T., 1998a. Volume, volatility, price, and profit when all traders are above average. Journal of Finance 53, 1887–1934.

Odean, T., 1998b. Are investors reluctant to realize their losses? Journal of Finance 53, 1775–1798.

Ofek, E., Richardson, M., 2003. DotCom mania: a survey of market efficiency in the Internet sector. Journal of Finance 58, 1113–1138.

Ofer, A., Melnik, A., 1978. Price deregulation in the brokerage industry: an empirical analysis. Bell Journal of Economics 9, 633–641.

Pastor, L., Stambaugh, R., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Pulvino, T., 1998. Do asset fire sales exist? An empirical investigation of commercial aircraft transactions. Journal of Finance 53, 939–978.

Ritter, J., 1991. The long-run performance of initial public offerings. Journal of Finance 42, 365–394.

Scheinkman, J., Xiong, W., 2003. Overconfidence and speculative bubbles. Journal of Political Economy, forthcoming.

Schultz, P., Zaman, M., 2001. Do the individuals closest to Internet firms believe they are overvalued? Journal of Financial Economics 59, 347–381.

Shefrin, H., Statman, M., 1985. The disposition to sell winners too early and ride losers too long: theory and evidence. Journal of Finance 40, 777–790.

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

Speiss, D., Affleck-Graves, J., 1995. Underperformance in long-run stock returns following seasoned equity offerings. Journal of Financial Economics 38, 243–267.

Stambaugh, R., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

Stein, J., 1995. Prices and trading volume in the housing market: a model with down-payment effects. Quarterly Journal of Economics 110, 379–406.

Stein, J., 1996. Rational capital budgeting in an irrational world. Journal of Business 69, 429–455.

Stigler, G., 1964. Public regulation of the securities markets. Journal of Business 37, 117–142.

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

Vayanos, D., 1998. Transaction costs and asset prices: a dynamic equilibrium model. Review of Financial Studies 11, 1–58.

Vuolteenaho, T., 2000. Understanding the aggregate book-to-market ratio and its implications to current equity-premium expectations. Unpublished working paper, Harvard University.