

Disagreement and the Stock Market

Harrison Hong and Jeremy C. Stein

Harrison Hong is Professor of Economics, Princeton University, Princeton, New Jersey. Jeremy C. Stein is Moise Y. Safra Professor of Economics, Harvard University, and Research Associate, National Bureau of Economic Research, both in Cambridge, Massachusetts. Their e-mail addresses are <hhong@princeton.edu> and <jeremy_stein@harvard.edu>, respectively.

Over the last twenty years, the field of behavioral finance has grown from a startup operation into a mature enterprise, with well-developed bodies of both theory and empirical evidence. On the empirical side, the benchmark null hypothesis is that one should not be able to forecast a stock's return with anything other than measures of its riskiness, such as its beta; this hypothesis embodies the familiar idea that any other form of predictability would represent a profitable trading rule and hence a free lunch to investors. Yet in a striking rejection of this null, a large catalog of variables with no apparent connection to risk have been shown to forecast stock returns, both in the time series and the cross-section. Many of these results have been replicated in a variety of samples and have stood up sufficiently well that they are generally considered to be established facts.

One prominent set of patterns from the cross-section has to do with medium-term momentum and post-earnings drift in returns. These describe the tendency for stocks that have had unusually high past returns or good earnings news to continue to deliver relatively strong returns over the subsequent six to twelve months (and vice-versa for stocks with low past returns or bad earnings news). Early work in this area includes Jegadeesh and Titman (1993) on momentum and Bernard and Thomas (1989, 1990) on post-earnings drift. Another well-established pattern is longer-run fundamental reversion—the tendency for “glamour” stocks with high ratios of market value to earnings, cashflows or book value to deliver weak returns over the subsequent several years (and vice-versa for “value” stocks with low ratios of market value to fundamentals). Standard references for this value-glamour phenomenon include Fama and French (1992) and Lakonishok, Shleifer and Vishny (1994).

On the theory side, research has proceeded along two distinct fronts. First, one needs to explain what prevents rational arbitrageurs from eliminating these and other predictable patterns in returns. Work in this “limits to arbitrage” vein has focused on the risks and market frictions that arbitrageurs face. These include simple transactional impediments, like short-selling constraints, as well as a variety of other complications. Potential arbitrageurs face the risk that when they bet against a given mispricing, this mispricing may subsequently worsen, with the ultimate correction coming only much later (DeLong et al., 1990; Shleifer and Summers, 1990). This risk is exacerbated by the fact that many of the most sophisticated would-be arbitrageurs are professional asset managers, who act as agents when they invest other people’s money. A professional manager has to worry that poor short-run performance will lead to withdrawals from his fund, leaving him liquidity constrained and unable to hang on to even those positions that he views as sure winners (Shleifer and Vishny, 1997).

Although research on limits to arbitrage is far from played out, it is fair to say that a broad consensus is emerging with respect to the key ideas and modeling ingredients. This should not be too surprising, given that the relevant tools all come from neoclassical microeconomics: arbitrageurs can be modeled as fully rational, with no need to appeal to any behavioral or psychological biases. For example, the work on the liquidity constraints associated with delegated arbitrage can be thought of as embedding familiar theories from corporate finance into an asset-pricing framework.

Much less consensus has been achieved on the second front, that which seeks to explain the *specific nature* of the patterns of predictability. Even taking as given that rational arbitrage cannot correct all instances of mispricing, what is it about the behavior

of other, presumably less rational investors that makes stock prices appear to underreact to certain types of information in the short run, but overreact in the longer run? Here it is by definition harder to proceed without entertaining deviations from the standard rational-agent paradigm, which necessarily opens up a can of worms. Different authors have taken very different approaches, including representative-agent models with rational beliefs but unconventional preferences, such as those associated with prospect theory (Barberis and Huang, 2001); representative-agent models with standard preferences but biased beliefs (Barberis, Shleifer and Vishny, 1998; Daniel, Hirshleifer and Subrahmanyam, 1998); and a variety of heterogeneous-agent models.

Excellent surveys of recent work in behavioral asset pricing include Hirshleifer (2001) and Barberis and Thaler (2003). In this paper, we do not attempt to be either as balanced or as comprehensive as these authors. Rather, we adopt the role of advocates, and argue in favor of one particular class of heterogeneous-agent models, which we call “disagreement” models. This category is fairly broad, encompassing work that has focused on the following underlying mechanisms: i) gradual information flow; ii) limited attention; and iii) heterogeneous priors. While these three mechanisms each have their own distinct features, both theoretically and in terms of empirical content, we argue below that they share important common elements. In particular, this class of models is at its heart about the importance of differences in the beliefs of investors.¹

Disagreement models have a number of attractive features. In our view, the most compelling is that they allow us to speak directly to the *joint behavior of stock prices and*

¹ So long as arbitrage by fully rational agents is limited, we do not need to assume that these disagreement mechanisms apply to all investors. It is sufficient that they apply to a significant subset of the universe, which may include both some individual investors as well as some professional money managers.

trading volume. Indeed, we find it hard to imagine a fully satisfying asset-pricing model—in either the rational or behavioral genres—that does not give a front-and-center role to volume. Trading volume is extremely large across virtually all developed stock markets, and many of the most interesting patterns in prices and returns are tightly linked to movements in volume. For example, high-priced glamour stocks tend to have higher volume than low-priced value stocks, all else equal. Also, controlling for a stock's ratio of price to fundamentals, its future returns tend to be lower when it has higher trading volume—in other words, trading volume appears to be an indicator of sentiment. In what follows, we argue that disagreement models offer a natural framework for understanding these and other related phenomena.

The Importance of Trading Volume

To help motivate our argument, we begin with some basic facts about the raw magnitude of trading volume. We then document the pervasive tendency for higher volume to accompany higher price levels, both in the time series and the cross section.

In 2005, the dollar value of trading volume on the New York Stock Exchange (NYSE) was \$14.1 trillion; on NASDAQ, it was \$10.1 trillion; on the London Stock Exchange, it was \$5.7 trillion; and on the Tokyo Exchange, \$4.5 trillion. Worldwide, the roughly 50 members of the World Federation of Exchanges accounted for a total of \$51.0 trillion of trading volume. In recent years, turnover on the NYSE has averaged about 100

percent (it was 102 percent in 2005), meaning that the entire market value of a typical firm changes hands about once a year.²

In conventional rational asset-pricing models with common priors—even those that allow for asymmetries in information across traders (Grossman and Stiglitz, 1980; Kyle, 1985)—the volume of trade is approximately pinned down by the unanticipated liquidity and portfolio rebalancing needs of investors. However, these motives would seem to be far too small to account for the tens of trillions of dollars of trade observed in the real world. This dissonance has led even the most ardent defenders of the traditional pricing models to acknowledge that the bulk of volume must come from something else—for example, differences in prior beliefs that lead traders to disagree about the value of a stock even when they have access to the same information sets.

Nevertheless, the implicit view taken by the traditionalists is that while such disagreement generates trading activity, these trades are idiosyncratic and therefore cancel each other out, with no consequences for prices. This implies that stock prices can continue to be analyzed in the usual representative-agent efficient-markets setting, with trading volume being left as a separate and effectively unconnected area of inquiry. Reflecting this “decoupling” point of view, many of the important first-generation papers on disagreement in financial markets focus on generating predictions for trading volume, but do not attempt to speak directly to questions of pricing (Varian, 1989; Harris and Raviv, 1993; Kandel and Pearson, 1995).

In contrast to this view, many early economists believed that elevated trading volume could play a central role in generating speculative bubbles. Kindleberger and

² Source: World Federation of Exchanges website, www.world-exchanges.org, and NYSE Fact Book, available at www.nyse.com.

Aliber (2005) point out that classical ideas about bubbles by Adam Smith, John Stuart Mill, Knut Wicksell and Irving Fischer were based on the concept of “overtrading,” the process whereby euphoric investors purchase shares solely in anticipation of future capital gains. Perhaps the classical economists were influenced by episodes such as the South Sea bubble of 1720. Carlos, Neal and Wandschneider (2006) document dramatic increases in turnover in the shares of the Bank of England, the East India Company, and the Royal African Company during that bubble. Their most detailed data pertain to trading in Bank of England stock. In the three years prior to the bubble there were approximately 2,000 transactions per year in the Bank’s stock, while in 1720—the year of the South Sea bubble—nearly 7,000 transactions occurred.

In a similar vein, accounts of the stock-market boom of the late 1920s such as Galbraith (1979) often emphasize the heightened level of trading volume in 1928 and 1929 as an important element of market dynamics. Here’s one way to get a sense for the intensity of trading activity during this period. Since 1900, NYSE share volume has set an all-time record on 74 days. Of these 74 record-breaking days, 10 were in 1928, and three were in 1929.³ After 1929, a new record was not set until April 1, 1968, following Lyndon Johnson’s announcement that he would not seek re-election.

Ofek and Richardson’s (2003) analysis of the recent internet bubble touches on the same themes, but with the benefit of much more complete stock-level data on trading volume. Figure 1 reproduces one of their main findings. The figure plots the average monthly level of turnover and prices, for both an index of internet stocks, as well as for the remainder of the non-internet stocks, over the period January 1997 to December

³ Source: Authors’ tabulations using data on NYSE daily share volume in NYSE listed issues available at www.nyse.com.

2002. As can be seen, the remarkable run-up in internet stock prices was accompanied by an extraordinary explosion in trading volume. *Monthly* turnover in internet stocks exceeds 50 percent in twelve out of the 24 months preceding the internet index's price peak in February 2000, with monthly turnover reaching as high as 101 percent in December 1998—an annualized rate of turnover of over 1,200 percent. By contrast, monthly turnover for non-internet stocks is generally in the 10-15 percent range over the same period, and only once does it creep slightly above 20 percent.⁴

The common thread across these episodes is that trading volume appears to act as an indicator of investor sentiment. In other words, when prices look to be high relative to fundamental values, volume is abnormally high as well. This basic relationship turns out to be quite general, arising not only in dramatic bubble-like situations, but also in broader cross-sectional and time-series samples.

Figure 2 illustrates the cross-sectional phenomenon. Each quarter in the period 1986 to 2005, we begin with the universe of the 1,000 largest stocks in the Center for Research in Security Prices (CRSP) database. We then compute exchange-adjusted monthly turnover for each stock, defined as that stock's turnover minus the average turnover of all stocks listed on the same exchange (either NYSE/AMEX or NASDAQ). Finally, we calculate average adjusted monthly turnover for each of two portfolios: i) a portfolio of high-priced glamour stocks—those ranked in the upper quintile of the universe by market-to-book ratio; and ii) a portfolio of low-priced value stocks—those in the lower quintile of the universe by market-to-book.

⁴ A caveat is that measured turnover in NASDAQ stocks tends to be a bit higher than in NYSE/AMEX stocks because of the dealer nature of the NASDAQ market. This difference could affect our comparisons because internet firms disproportionately trade on the NASDAQ. However, if we adjust the data to account for exchange-wide mean levels of turnover, the results are very similar to those in Figure 1.

As Figure 2 shows, glamour stocks tend to have significantly higher turnover than value stocks (see also Piqueira, 2006). This differential is largest during the run-up of the internet bubble, but it is apparent in virtually every month in the sample. In particular, glamour stocks continue to display more turnover even in 2000-2001, when the internet bubble is collapsing and the returns to glamour stocks fall far below those of value stocks. So the phenomenon captured in Figure 2 reflects more than just the so-called “disposition effect” (Odean, 1998), whereby investors tend to be reluctant to sell stocks that have recently declined in value.

Figure 3 considers the relationship between trading volume and prices in the time-series dimension. Here we go back to 1901, and plot for each year the annual real return on the Standard and Poor’s (S&P) 500, along with the percentage change in turnover on the NYSE from the preceding year. Thus we are now looking at how *changes* in price levels covary with *changes* in the level of trading volume; one advantage of this differencing approach is that it removes low-frequency time trends from the turnover data. The resulting relationship is visually striking, and is also confirmed statistically: the correlation between the two series is a highly significant 0.49.

The connections between prices and trading volume documented in this section should serve to whet the reader’s appetite for asset-pricing theories in which volume plays a central role. However, we have not yet addressed two key questions. First, what are the underlying mechanisms, either at the level of market structure or individual cognition, that give rise to disagreement among traders and hence to trading volume? Second, how do these mechanisms simultaneously generate mispricings of one sort or another? Or said differently, why do they not simply lead to trades that cancel each

other out in terms of price effects, as implicitly assumed by the traditional model? The next two sections seek to answer these questions.

Before proceeding, however, we should clarify the distinctions between our approach and the well-known model of DeLong et al. (1990). In DeLong et al., a group of smart-money arbitrageurs interacts with a group of noise traders who are subject to exogenous sentiment shocks. This model also generates trading volume, to the extent that the valuations of the noise traders shift relative to those of the arbitrageurs—that is, to the extent that equilibrium prices move closer to or further away from fundamental values. At the same time, the amount of volume generated by the DeLong et al. model is not always large. This is most clearly seen by considering a scenario in which the realization of the sentiment shock is close to zero. In this case, the level of prices remains unchanged relative to fundamentals, and trading volume is negligible.

In contrast, our approach can be thought of as one in which we start by looking for a disagreement mechanism that can *robustly explain* the large observed levels of trading volume—i.e., one that can generate a lot of trading activity even when prices are not moving relative to fundamentals. Having identified such a disagreement mechanism, the next step is to ask whether and how it can influence prices in addition to volume. In so doing, we hope to both endogenize the degree of investor sentiment that DeLong et al. (1990) take as exogenous, as well as to derive a number of further testable implications. For example, we argue below that the heterogeneous-priors mechanism, when combined with short-selling constraints, makes the prediction that an increase in the number of news stories about a company has a systematic tendency to drive prices up—that is, to endogenously raise the level of effective sentiment in the market.

Mechanisms That Can Generate Investor Disagreement

Gradual Information Flow

Gradual information flow is an important feature of asset markets (Hong and Stein, 1999). As a result of either the technology of information distribution, or investor segmentation and specialization, certain pieces of value-relevant information will arrive in the hands of some investors before others. If the information is positive, those who receive it first will revise their valuations of the stock in question upward, while those who have not yet seen it will have unchanged valuations. As a result, disagreement between investors in the two groups will increase, and those in the former group will tend to buy from those in the latter.

One of the cleanest examples of this phenomenon comes from Huberman and Regev (2001), who examine the stock-market behavior of a single biotechnology firm, EntreMed. On May 3, 1998, the Sunday *New York Times* carried a front-page story on recent innovations in cancer research, and featured EntreMed prominently. The next day, Monday May 4, EntreMed stock, which had closed the previous Friday at \$12/share, shot up on heavy volume, ending the day at \$52/share.

What is remarkable about this episode is that the front-page *Times* story contained essentially no real news: the substance of the story had been reported five months earlier, in November 1997, in the scientific journal *Nature*, as well as in the popular media (including the *Times* itself, in a less high-profile article on page A28). These earlier stories were also accompanied by positive bumps in both EntreMed's stock price and

trading volume, though neither was nearly as dramatic as the ones induced by the front-page *Times* story more than five months later. Moreover, while EntreMed subsequently retreated from its peak of \$52/share on May 4, 1998, it closed above \$30/share in the following three weeks. This suggests that, although there was an element of short-run overreaction, a large fraction of the impact of the front-page *Times* story was permanent.

A natural interpretation of these events is that there are two types of investors in EntreMed: a small group of scientific specialists who read publications like *Nature*; and a larger group of generalists who get their information from sources like the front page of the *Times*. Information flows in such a way that the specialists get certain pieces of news before the generalists; this leads to trading volume around the news-release events, and also to an apparent gradual response of prices to the substance of the news itself. That is, EntreMed's stock price did not go all the way up immediately when the *Nature* story was first released, but rather took several months to incorporate its implications fully.⁵

A key theoretical subtlety is that this information structure does not suffice to generate the above patterns in prices and volume, at least not on average. Gradual information flow by itself can be entirely consistent with a rational model with costs of information acquisition, in which specialists find it cheaper to acquire certain types of information before generalists. What is also required to get interesting price and volume effects is that, unlike in a classical rational-expectations setting, investors do not fully take into account the fact that they may be at an informational disadvantage, and hence

⁵ If one takes the view that specialists and generalists interpret news identically once they each see it, one needs to assume that the specialists were either small in number or very risk-averse to make the numbers in this case work out—if not, why would the specialists alone not have bid the price closer to its long-run equilibrium value? An alternative approach is to argue that at least some of the generalists interpreted the news more favorably than the specialists. This approach brings in elements of the heterogeneous-priors mechanism, which we discuss below. With the addition of a short-sales constraint, it may also do a better job of explaining the apparent short-run overreaction of the price to the front-page *Times* story.

do not draw the correct inferences from the trades of others. The failure of investors to update in this sophisticated way is often labeled “overconfidence” in the behavioral-finance literature, in the sense that investors overestimate the precision of their own information and hence underestimate the extent to which they should be learning from the trades of others. But it could equally well come from a simple lack of understanding about the structure of the environment.

In the context of the EntreMed example, this theory means that the generalists who do not read *Nature* tend to stick to their earlier (lower) valuations when the *Nature* story first comes out, even though the trades of the specialists lead to a potentially observable spike in demand. This lack of sophisticated updating on the part of the generalists leads both to a sluggish adjustment of prices to the *Nature* story, as well as to trade between the generalists and the better-informed specialists. In other words, where this model differs from a rational-expectations model is that the generalists do not deduce the information content of the *Nature* story by looking at subtle clues like the trading behavior of the specialists; rather, they only absorb this information when they are hit over the head with it, in the form of a front-page *New York Times* story.

A few recent papers provide broader evidence of an underreaction effect in stock prices very much in the spirit of the EntreMed example. For example, Menzly and Ozbas (2006) and Cohen and Frazzini (2006) find that when a particular firm or industry has positive stock returns over some interval, customers or suppliers of that firm or industry tend to experience positive stock returns over a subsequent interval, and vice-versa for negative returns. Hong, Torous and Valkanov (2006) offer a related set of findings. Thus, information appears to flow gradually across industries, perhaps because each industry

has its own set of specialist investors who focus on uncovering the most directly industry-relevant information, and who only slowly become aware of events in related industries.

Limited Attention

Several recent papers, including Hirshleifer and Teoh (2003) and Peng and Xiong (2006), stress the idea of limited attention, whereby cognitively-overloaded investors pay attention to only a subset of publicly-available information.⁶ For many practical purposes, this idea boils down to almost the same thing as gradual information flow, albeit with less emphasis on the dynamics of information diffusion. Also, as with gradual information flow, limited attention per se is not sufficient to generate interesting patterns in prices or volume. Rather, limited attention needs to be combined with the assumption that investors are also unsophisticated in a second, logically distinct way: when trading with others, they do not adjust for the fact that they are basing their valuations on only a subset of the relevant information.

Although the differences between limited attention and gradual information flow may be somewhat semantic, the limited-attention label does hint at some interesting further nuances. In particular, a focus on limited attention suggests that, holding fixed the substantive content associated with a news release, the response of prices and trading volume to the release will be larger when it is broadcast in an “attention-grabbing” manner. This insight implies a potentially important role for the media in shaping the behavior of the stock market, as front-page *New York Times* coverage in the EntreMed

⁶ Again, with limited arbitrage, it is not necessary to assume that *all* investors are cognitively-overloaded; those who are can be thought of as either individuals or institutions. In Hong, Stein and Yu (2006), something like limited attention emerges endogenously, as investors search for simple forecasting models that allow them to economize on the cognitive costs of information processing.

case illustrates. Similarly, Klibanoff, Lamont and Wizman (1998) document that the prices of closed-end country funds respond more strongly to changes in the funds' net asset values when the country in question is also featured in a front-page story in the *New York Times*.

The converse proposition is that a news release will have less effect if investors with limited attention are distracted for some reason. DellaVigna and Pollet (2006) put forward a clever piece of evidence in support of this hypothesis. They find that when a firm announces its earnings on a Friday, the resulting volume is less than for those announcements that fall on other days of the week, and the stock price underreacts by more. Their interpretation is that investors become distracted over the weekend, and partially forget about the implications of the news by the time they have a chance to act on it the next Monday morning.

Heterogeneous Priors

Even if a given piece of news is made publicly available to all investors simultaneously, and even if they all pay attention to it, the news can nevertheless increase their disagreement about the fundamental value of the stock in question. As Harris and Raviv (1993) and Kandel and Pearson (1995) discuss, this outcome will occur if investors have different economic models that lead them to interpret the news differently.

To take a specific example, suppose that a firm announces that its earnings are up by 10 percent from the previous quarter. To an investor who had expected no increase in earnings, and who thinks that earnings shocks are permanent, this announcement might imply an increase in the present value of expected future earnings of roughly 10 percent

as well. To a second investor who had also expected no increase in earnings, but who thinks that earnings shocks are relatively transitory, the news would again be seen as positive, though less so than for the first investor. Finally, for a third investor who had expected a 20 percent earnings increase, the news would be a disappointment, and would lead to a downward revision in expected future earnings.

This example suggests that even when the three investors all observe the same earnings announcement, they may be induced to trade with one another. Again, however, in order to derive predictions for trading volume, one needs to combine heterogeneous priors with an assumption that the investors do not fully update their beliefs based on each other's trading decisions—that is, the investors agree to disagree in equilibrium.

The behavior of trading volume in the wake of public earnings announcements provides a striking illustration of the heterogeneous-priors mechanism. Figure 4 plots abnormal daily turnover (defined for each stock as its daily turnover minus its average turnover in the preceding 250 trading days) over an interval covering the 15 trading days before and after a quarterly earnings announcement; our sample is the 1,000 largest stocks in the CRSP database in each quarter, and the period is 1986-2005. Turnover spikes up sharply when earnings news is released, and remains substantially elevated for more than a week afterwards. This pattern is precisely the opposite of what one would expect based on a simple rational-expectations model with common priors, where public information should have the effect of reducing disagreement, rather than increasing it.

Figure 4 also suggests another potential channel through which media coverage can matter for the stock market: if one thinks of the arrival of public news as creating the raw fodder for disagreement, then increases in the intensity of media coverage can act as

a direct stimulus to trading. During the internet bubble, for example, disproportionately more media attention was paid to internet stocks than to non-internet stocks (Bhattacharya et al., 2004). We argue below that this excess media coverage may help to explain not only the extraordinary levels of trading volume in these stocks (as shown in Figure 1), but also to some extent their elevated prices.

Disagreement, Prices and Volume: Two Applications

The primary appeal of the disagreement mechanisms described in the previous section is that they can be embedded in models that both: i) provide an explanation for well-known asset pricing patterns, such as medium-term momentum and the overpricing of glamour stocks; and ii) deliver additional testable restrictions having to do with the joint behavior of prices and volume. We now offer two illustrations of this point.

Medium-term momentum

Consider a simple model of the momentum effect first documented by Jegadeesh and Titman (1993). There is a stock that will pay a liquidating dividend at time 2 of $D = A + B$, where A and B are two independent mean-zero normal random variables that can be thought of as the realizations of two distinct pieces of public information. At some initial time 0, A and B are unknown, so the time-0 expectation of D is simply zero. Assume for simplicity that the interest rate is zero, and that there is also no risk premium for holding the stock (an assumption which is justified if the supply of the stock is

negligibly small relative to investors' aggregate risk tolerance). It then follows that the time-0 price, denoted by P_0 , is also equal to zero.

Now suppose that when the realizations of A and B become available at time 1, a fraction f of the investing public only sees A , and the remaining fraction $(1 - f)$ only sees B . Moreover, those that see A believe that A is all that matters for forecasting D (they mistakenly believe that $D = A$) and analogously for those that see B . These assumptions capture in a reduced-form way the essence of both the gradual-information-flow and limited-attention stories. If investors can frictionlessly take either long or short positions, and given the vanishingly small supply of the stock, the assumptions imply that the market price at time 1, P_1 , is equal to $fA + (1 - f)B$: simply put, the price is nothing more than the weighted average of investors' expectations of D at this time. The dollar return on the stock from time 0 to time 1, denoted by $R_1 \equiv (P_1 - P_0)$, is therefore also equal to $fA + (1 - f)B$. And since the stock pays a liquidating dividend of D at time 2, the return from time 1 to time 2, denoted by R_2 , is just $(D - P_1)$, which is equal to $(1 - f)A + fB$.

With these definitions in hand, one can pose two questions. First, is there positive momentum in returns from time 1 to time 2? And second, how does the degree of momentum relate to trading volume? In the current context, momentum can be summarized by the covariance (or equivalently, the correlation) of the time-1 and time-2 returns, R_1 and R_2 . That is, there is momentum when a positive (negative) return at time 1 tends to be followed by another positive (negative) return at time 2.

It is straightforward to show that momentum in stock returns is generally positive in this model. Moreover, the degree of momentum is maximized when $f = (1 - f) = 1/2$,

that is, when fraction of investors attending to each piece of information is the same.⁷ The intuition for this result is as follows. If, say, nobody pays attention to A ($f = 0$), the market completely underreacts to A at time 1, and the price response to A is therefore entirely delayed until time 2. However, there is no predictability of the time-2 return based on this delayed response, since the time-1 return contains no information about A . In contrast, with $f = \frac{1}{2}$, half of the information about A gets into the price at time 1, and the other half gets into the price at time 2, which maximizes the momentum effect.

Interestingly, the model also implies that the magnitude of the momentum effect will be increasing with average trading volume. Clearly, if all investors focus only on A or only on B at time 1 ($f = 0$ or $f = 1$), then there is never any disagreement among them, and hence no trading volume. In contrast, when half of the investing population focuses on each of the two variables ($f = \frac{1}{2}$), expected volume is maximized.⁸ Thus both momentum and expected trading volume increase together as the shares of investors attending to A and B move toward equality.

The bottom line from this exercise is that not only can a model based on gradual information flow or limited attention shed light on the momentum phenomenon, it also

⁷ As noted above, momentum can be measured by the covariance of R_1 and R_2 . Direct calculation based on the formulas for R_1 and R_2 in the text yields:

$$\text{cov}(R_1, R_2) = f(1 - f)(v_A + v_B),$$

where v_A and v_B are the variances of A and B respectively. It follows immediately that momentum is generally positive, and that momentum is maximized when $f = (1 - f) = \frac{1}{2}$.

⁸ To see this formally, observe that volume is proportional to the absolute deviation between the A -investors' valuations and the market price, multiplied by the number of A -investors. This quantity is given by $f|A - P_1|$. Substituting in the value of P_1 , and taking expectations, expected volume satisfies:

$$\text{Expected Volume} \approx f(1 - f)E(|A - B|).$$

Therefore, expected volume is maximized when $f = (1 - f) = \frac{1}{2}$.

suggests an additional testable hypothesis: the momentum effect should be most pronounced, all else equal, in those stocks for which average trading volume is high. As it happens, this prediction receives some support in empirical work by Lee and Swaminathan (2000). They document that momentum trading strategies of the sort identified by Jegadeesh and Titman (1993)—in which one buys past-winner stocks and sells past-loser stocks—perform significantly better when restricted to subsamples of stocks with high recent turnover.⁹

A final point to note about this model is that, in a qualitative sense, it is consistent with much higher levels of volume than the noise-trader model of DeLong et al. (1990). Recall that in DeLong et al., volume falls to zero in those periods when price is not moving relative to fundamental value. Here, by contrast, there can be substantial volume in a given stock even over an interval when its price is unchanged. To see this, suppose that $f = \frac{1}{2}$, and that the realizations of A and B at time 1 are equally large in absolute value but of the opposite sign—say +100 and –100 respectively. In this case, the time-1 price will be equal to both the time-0 price and the long-run fundamental value: all are zero. But the pronounced divergence in the valuations of the A -investors and the B -investors will lead them to trade aggressively with one another at time 1. The key distinction relative to DeLong et al. is that here we have two sets of less-than-fully rational investors who each make different mistakes. Even when these mistakes cancel out in terms of their effect on prices, they can still generate a large volume of trade.

⁹ While their results are suggestive, we would not want to construe Lee and Swaminathan (2000) as a decisive test of the model sketched above. There are many reasons other than those in our model why volume may vary over time and across stocks, including differences in trading costs. To provide a sharper test of the hypothesis, one would ideally like to isolate variation in volume that is orthogonal to these other factors, and that is more likely to reflect the sort of disagreement that our model highlights. Lee and Swaminathan do not do this—they just divide stocks into groups based on raw levels of turnover.

Overpriced Glamour Stocks

We turn next to the phenomenon of overpriced glamour stocks, perhaps best exemplified by internet firms during the bubble period from 1998 to 2000. The story here also centers on disagreement, though it is probably most convenient to think of the disagreement as coming from the heterogeneous-priors mechanism—whereby different investors may reach different conclusions even when exposed to the same public news.

In addition to disagreement, we now also add a second key ingredient to the mix: short-sales constraints. In particular, we assume that investors either cannot or will not short-sell stocks. Thus if an investor thinks that a given stock is overvalued, the investor does not sell the stock short, but rather just sits out of the market. Empirically, this assumption seems well-founded. Individual investors and many types of institutions rarely take short positions. (Almazan et al. 2004 document the virtual absence of shorting by mutual funds.) While hedge funds are known for shorting more aggressively, the total number of shares sold short at any given time remains a small fraction of shares outstanding (Lamont and Stein, 2004). Moreover, our theoretical predictions below do not depend on literally all investors being short-sales constrained—it is sufficient that such constrained investors represent a meaningful fraction of the market’s risk tolerance.

The interaction of disagreement and short-sales constraints has been analyzed in both static and dynamic models. The static models originate in the work of Miller (1977), who points out that in the presence of a short-sales constraint, the valuations of optimists will be reflected in a stock’s price, but the valuations of pessimists will not. Thus, even if the valuation of the average investor is unbiased, the stock price will be

biased upward. In contrast, in a world without constraints on short-selling, the stock price (loosely speaking) reflects the value-weighted average opinion.

One interesting result that emerges from Miller's (1977) model is that, holding fixed the average investor's valuation, the degree of overpricing increases as the dispersion of valuations rises—that is, as disagreement becomes more pronounced. Using different proxies for disagreement, both Diether, Malloy and Scherbina (2002) and Chen, Hong and Stein (2002) develop evidence that supports this prediction of Miller's model. Again, the intuition is that market prices are driven by the optimists, so if the optimists become more optimistic, prices must go up, even if at the same time the pessimists become more pessimistic.

At first glance, it is tempting to map this result into a statement about trading volume and overpricing. However, taken literally, the Miller (1977) model cannot speak to volume, because it is a static model in which investors take initial positions in the stock and never rebalance their positions before the stock liquidates. In other words, trading volume in any given interval does not come from the existing level of disagreement, but rather from changes in the level of disagreement; this idea is difficult to capture in a static setting.

Dynamic models with disagreement and shorting constraints are thus better suited to studying the joint behavior of volume and overpricing. This line of work was initiated by Harrison and Kreps (1978); recent contributions include Scheinkman and Xiong (2003) and Hong, Scheinkman and Xiong (2006). In these models, investors continually update their valuations based on their personal interpretations of incoming news, and

trade between any two investors occurs whenever their valuations “cross”—that is, whenever the more optimistic of the two switches to being the more pessimistic.

A central prediction of these dynamic models is that a positive correlation exists between trading volume and the degree of overpricing.¹⁰ To see the logic, it is useful to think of two otherwise similar stocks, with the one difference being that stock i is subject to a greater arrival rate of public news than stock j . This difference could arise either because there really is more fundamental uncertainty about stock i (perhaps it represents a newer and more untested technology) or because stock i is a media darling and attracts more coverage per unit of real information. In a world in which investors interpret news differently, the greater news stimulus associated with stock i will lead to more time-series variance in investors’ relative valuations, and hence to more trading volume via the “crossing” effect outlined above; Scheinkman and Xiong (2003) demonstrate this point formally. Moreover, as a result of the short-sales constraint, stock i will be more overpriced than stock j , because the greater news stimulus creates more disagreement and hence and set of more extreme optimists.

This prediction conforms well with the patterns seen in Figures 1-3, all of which suggest that trading volume is greater in higher-priced stocks. But since the theoretical prediction is more precisely about the relationship between volume and *overpricing*, a much sharper test of the theory would be to ask: controlling for a stock’s price level

¹⁰ The dynamic models are also capable of producing a greater degree of overpricing than the static ones. In Miller (1977), the price can never exceed the valuation of the most optimistic investor, whereas in Harrison and Kreps (1978), it can. This is because of a resale option that arises in the dynamic setting: an optimist may be willing to pay a price that exceeds his own current valuation of the stock, on the premise that he will be able to resell it to somebody else next period, if that other person revises his opinion and becomes even more bullish on the stock.

(perhaps measured by its market-to-book ratio), does higher volume forecast lower future stock returns?

Several studies suggest that the answer to this question is “yes,” both in the time series and the cross section. In the time series, Baker and Stein (2004) document a negative relationship between turnover on the NYSE and the subsequent year’s return on the aggregate market over the period 1933-1999, even controlling for the market-wide dividend-to-price ratio. This relationship is especially strong when forecasting an equal-weighted (as opposed to value-weighted) market index. In this case, a one-standard deviation increase in detrended turnover—equivalent to annual turnover going up by 12.3 percentage points as against a sample mean of 33.8 percent—is associated with a roughly 10-percentage-point reduction in the next year’s expected returns.

Such time-series return-forecasting exercises are by their nature statistically fragile. More statistically robust support for the hypothesis comes from cross-section section, where a large number of papers, including Datar et al. (1998) and Brennan et al. (1998), have found that measures of trading activity reliably forecast lower returns, controlling for other well-known cross-sectional predictors such as market-to-book, momentum and firm size.¹¹ However, a possible alternative interpretation of these results is that stocks vary in their liquidity-related trading costs. All else equal, a stock with lower trading costs should be expected to have higher turnover, as well as lower returns, via a purely rational liquidity-premium effect.

¹¹ The theory is a bit hazy as to whether one should ideally try to forecast returns based on relatively permanent cross-stock differences in volume, as opposed to more transitory within-stock differences. On the one hand, purging the permanent cross-stock component may help to control for differences in other characteristics of stocks, such as trading costs. On the other hand, some elements of the theory sketched above, such as the intensity of media coverage, may also have quite persistent cross-stock variation.

Piqueira (2006) attempts to separate the two stories. She controls carefully for market liquidity, using bid-ask spreads at the stock level, as well as the price impact of trade.¹² She still finds that turnover has a significant negative association with future returns, in a sample running from 1993-2002. For example, among NASDAQ stocks, mean monthly turnover in the sample period is 15.0 percent. If monthly turnover goes up by one standard deviation, or 18.9 percentage points, monthly returns are forecast to fall by 0.75 percent, or roughly 9.0 percent on an annualized basis.

Another interesting tidbit comes from Lamont and Thaler's (2003) study of mispricing in tech-stock carve-outs. They find that when the law of one price is violated, the higher-priced security has trading volume that is many times that of the lower-priced security. For example, in the well-known Palm/3Com case, turnover in shares of Palm (the overpriced subsidiary) was vastly higher than turnover in shares of 3Com (the underpriced parent firm).

Finally, Frazzini and Lamont (2006) examine the returns to stocks around their earnings-announcement dates. Recall from Figure 4 that trading volume is sharply elevated in the days surrounding earnings announcements. Frazzini and Lamont further show that stock returns are on average abnormally positive over this same event-time interval. This result is on its face surprising, because it involves averaging over those announcements that convey good news and those that convey bad news. Moreover, the on-average-positive effect is strongest for those stocks that tend to experience the biggest proportional surge in volume around the earnings date.

¹² Following Kyle (1985), the price impact of trade is roughly the amount by which a buy (sell) order of a given size leads to an increase (decrease) in a stock's price—i.e., it is the derivative of price with respect to order flow. Higher values of this measure reflect lower levels of liquidity in the stock in question.

One way to interpret these results, and to connect them to the others discussed above, is to think that once every quarter, each stock goes through a week of unusually high news exposure associated with its earnings release. This news stimulus may spark increased disagreement among those investors who were already following the stock, and is also likely to grab the attention of those who were not. In either case, the end result is both more trading volume and—in the presence of a short-sales constraint—concurrent upwards pressure on the price. It is as if for a week out of the quarter, each stock experiences an exogenous impulse that makes it a little bit more like a glamour stock than it normally is, with the predicted consequences for both prices and volume.

Conclusion

The enduring appeal of classical asset-pricing theory over the last several decades owes much to its success in forging a consensus around a foundational modeling platform. This platform consists of a core set of assumptions that have been widely-accepted by researchers working in the field as reasonable first-order descriptions of investor behavior, and that—just as importantly—lend themselves to elegant, powerful and tractable theorizing.

If behavioral finance is ever to approach the stature of classical asset pricing, it will have to move beyond being a large collection of empirical facts and competing one-off models, and ultimately reach a similar sort of consensus. While this goal seems well within sight in the part of the field that explores limits to arbitrage, it is much further away in the part that seeks to understand the origins of market mispricings. Many horses

are still running in this latter race, and it is not at all clear whether a decisive winner will emerge in the foreseeable future.

Nevertheless, our view is that, taken collectively, the disagreement models described above represent the best horse on which to bet.¹³ Disagreement models uniquely hold the promise of being able to deliver a comprehensive joint account of stock prices and trading volume, which we consider to be one of the highest priorities for theoretical work in asset pricing. Moreover, the modeling ingredients are highly tractable, for two reasons. First, preferences are taken to be completely standard—so the modeler can work with constant absolute risk aversion utility or whatever other functional form makes life easiest. Second, in disagreement models, investors' beliefs are often a simple function of just their own priors and the signals that they each observe directly; this is in contrast to rational-expectations models, where each investor must also update based on inferences about others' priors and signals. As a result, disagreement models can be usefully deployed in everything from simple illustrative examples (like the two-period model of momentum discussed earlier in this paper) to elegant continuous-time formulations (like Scheinkman and Xiong, 2003). We hope that future research with disagreement models continues to develop their potential.

¹³ Interestingly, models of disagreement are also beginning to attract more attention in other fields. One example from macroeconomics is the work of Mankiw, Reis and Wolfers (2003), who study disagreement about inflation expectations.

Acknowledgements

We are grateful to the National Science Foundation for financial support, to Sam Hanson for outstanding research assistance, and to Malcolm Baker, Stefano DellaVigna, James Hines, Owen Lamont, Oguzhan Ozbas, Andrei Shleifer, Timothy Taylor, Hans-Joachim Voth, Michael Waldman and Jialin Yu for helpful comments.

REFERENCES

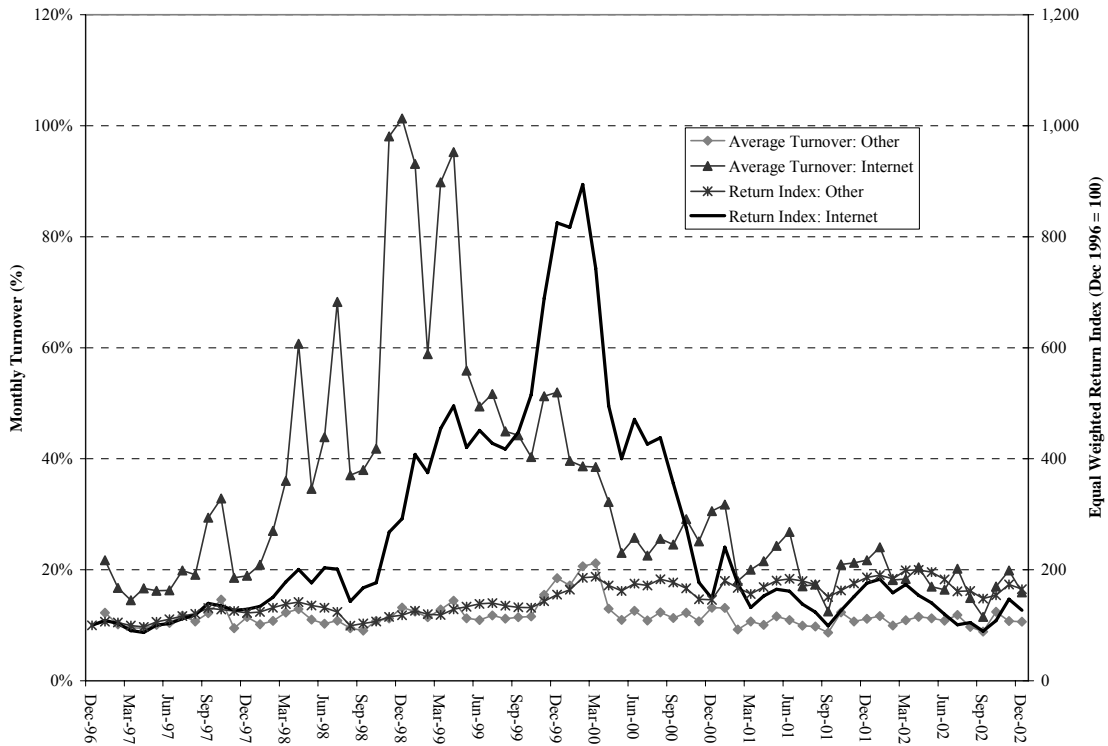
- Almazan, Andres, Keith C. Brown, Murray Carlson, and David A. Chapman, 2004, Why constrain your mutual fund manager? *Journal of Financial Economics* 73, 289-321.
- Baker, Malcolm, and Jeremy C. Stein, 2004, Market liquidity as a sentiment indicator, *Journal of Financial Markets* 7, 271-299.
- Barberis, Nicholas, and Ming Huang, 2001, Mental accounting, loss aversion and individual stock returns, *Journal of Finance* 56, 1247-1292.
- Barberis, Nicholas, Andrei Shleifer, and Robert W. Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307-343.
- Barberis, Nicholas, and Richard Thaler, 2003, A survey of behavioral finance, in George Constantinides, Milton Harris and Rene Stulz, ed.: *Handbook of the Economics of Finance* (North-Holland).
- Bernard, Victor L., and J. Thomas, 1989, Post-earnings announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27, 1-48.
- Bernard, Victor L., and J. Thomas, 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.
- Bhattacharya, Uptal, Neal Galpin, Rina Ray, and Xiaoyun Yu, 2004, The role of the media in the internet IPO bubble, Indiana University Working Paper.
- Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam, 1998, Alternative factor specifications, security characteristics and the cross-sectional of expected returns, *Journal of Financial Economics* 49, 345-373.
- Carlos, Ann, Larry Neal, and Kirsten Wandschneider, 2006, Dissecting the anatomy of exchange alley: The dealings of stockjobbers during and after the South Sea bubble, University of Illinois Working Paper.
- Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171-205.
- Cohen, Lauren, and Andrea Frazzini, 2006, Economic links and predictable returns, Yale University Working Paper.

- Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and over-reactions, *Journal of Finance* 53, 1839-1885.
- Datar, Vinay, Narayan Naik, and Robert Radcliffe, 1998, Liquidity and stock returns: An alternative test, *Journal of Financial Markets* 1, 203-219.
- DellaVigna, Stefano, and Joshua Pollet, 2006, Investor inattention, firm reaction, and Friday earnings announcements, University of California at Berkeley Working Paper.
- De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703-738.
- Diether, Karl, Christopher Malloy, and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, *Journal of Finance* 52, 2113-2141.
- Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected returns, *Journal of Finance* 47, 427-465.
- Frazzini, Andrea, and Owen Lamont, 2006, The earnings announcement premium and trading volume, Yale University Working Paper.
- Galbraith, John Kenneth, 1979. *The Great Crash 1929*, 50th anniversary edition (Houghton Mifflin Company, Boston MA).
- Grossman, Sanford J., and Joseph Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393-408.
- Harris, Milton, and Arthur Raviv, 1993, Differences of opinion make a horse race, *Review of Financial Studies* 6, 473-506.
- Harrison, J. Michael, and David M. Kreps, 1978, Speculative investor behavior in a stock market with heterogeneous expectations, *Quarterly Journal of Economics* 93, 323-336.
- Hirshleifer, David, 2001, Investor psychology and asset pricing, *Journal of Finance* 56, 1533-1597.
- Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure and financial reporting, *Journal of Accounting and Economics* 36, 337-386.
- Hong, Harrison, Jose Scheinkman, and Wei Xiong, 2006, Asset Float and Speculative Bubbles, *Journal of Finance* 61, 1073-1117.
- Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading and overreaction in asset markets, *Journal of Finance* 54, 2143-2184.

- Hong, Harrison, Jeremy C. Stein, and Jialin Yu 2006, Simple forecasts and paradigm shifts, *Journal of Finance*, forthcoming.
- Hong, Harrison, Walter Torous, and Rossen Valkanov, 2006, Do industries lead stock markets?, *Journal of Financial Economics*, forthcoming.
- Huberman, Gur, and Tomer Regev, 2001, Contagious speculation and a cure for cancer: A non-event that made stock prices soar, *Journal of Finance* 56, 387-396.
- Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 93-130.
- Kandel, Eugene, and Neil D. Pearson, 1995, Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy* 103, 831-872.
- Kindleberger, Charles P., and Robert Z. Aliber, 2005, *Manias, Panics and Crashes: A History of Financial Crises*, 5th edition (John Wiley and Sons, Inc.: Hoboken, NJ).
- Klibanoff, Peter, Owen Lamont and Thierry Wizman, 1998, Investor reaction to salient news in closed-end country funds, *Journal of Finance* 53, 673-699.
- Kyle, Albert, 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1335.
- Lamont, Owen, and Jeremy C. Stein, 2004, Aggregate short interest and market valuations, *American Economic Review Papers and Proceedings* 94, 29-32.
- Lamont, Owen, and Richard Thaler, 2003, Can the market add and subtract? Mispricing in tech stock carve-outs, *Journal of Political Economy* 111, 227-268.
- Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation and risk, *Journal of Finance* 49, 1541-1578.
- Lee, Charles, and Bhaskaran Swaminathan, 2000, Price momentum and trading volume, *Journal of Finance* 55, 2017-2069.
- Mankiw, N. Gregory, Ricardo Reis and Justin Wolfers, 2003, Disagreement about inflation expectations, *NBER Macroeconomics Annual*, 209-248.
- Menzly, Lior, and Oguzhan Ozbas, 2006, Cross-industry momentum, USC Working Paper.
- Miller, Edward, 1977, Risk, uncertainty, and divergence of opinion, *Journal of Finance* 32, 1151-1168.

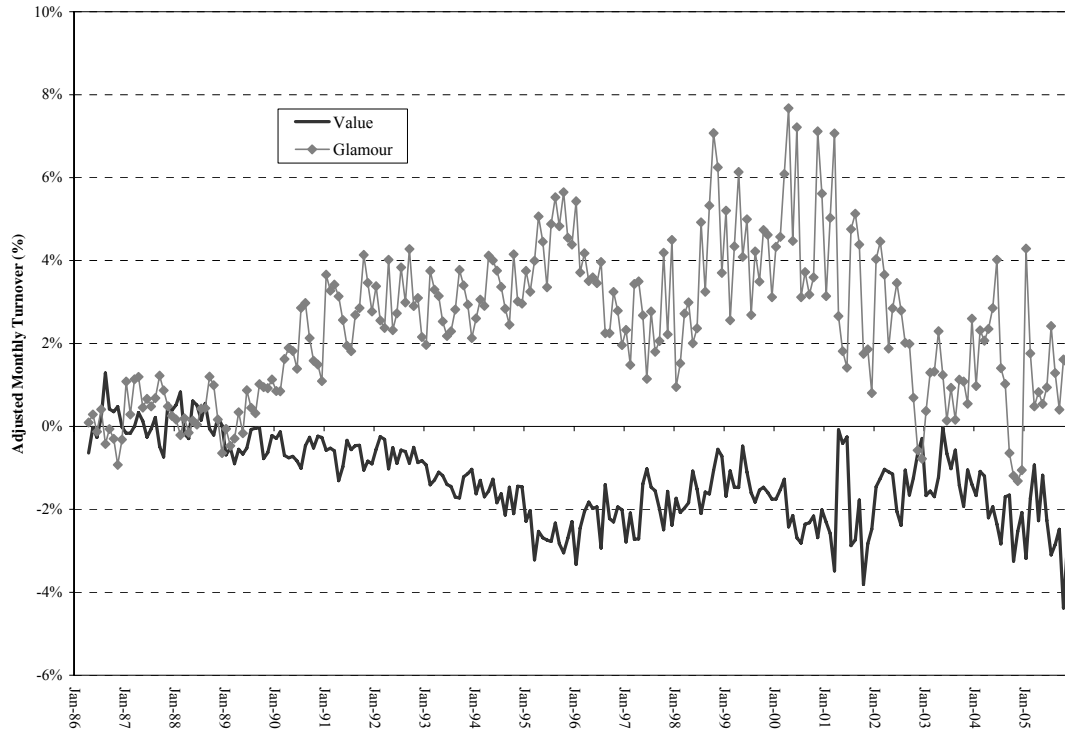
- Odean, Terrance, 1998, Are investors reluctant to realize their losses?, *Journal of Finance* 53, 1775-1798.
- Ofek, Eli, and Matthew Richardson, 2003, Dotcom mania: The rise and fall of internet stock prices, *Journal of Finance* 58, 1113-1137.
- Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics*, forthcoming.
- Piqueira, Natalia, 2006, Trading activity, illiquidity costs and stock returns, University of Houston Working Paper.
- Scheinkman, Jose, and Wei Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183-1219.
- Shleifer, Andrei, and Lawrence H. Summers, 1990, The noise trader approach to finance, *Journal of Economic Perspectives* 4, 19-33.
- Shleifer, Andrei, and Robert Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35-55.
- Varian, Hal R., 1989, Differences of opinion in financial markets, in CC. Stone, ed. *Financial Risk: Theory, Evidence, and Implications: Proceedings of the 11th Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis* (Kluwer Academic Publishers, Boston MA).

Figure 1
Prices and Turnover for Internet and Non-Internet Stocks, 1997-2002



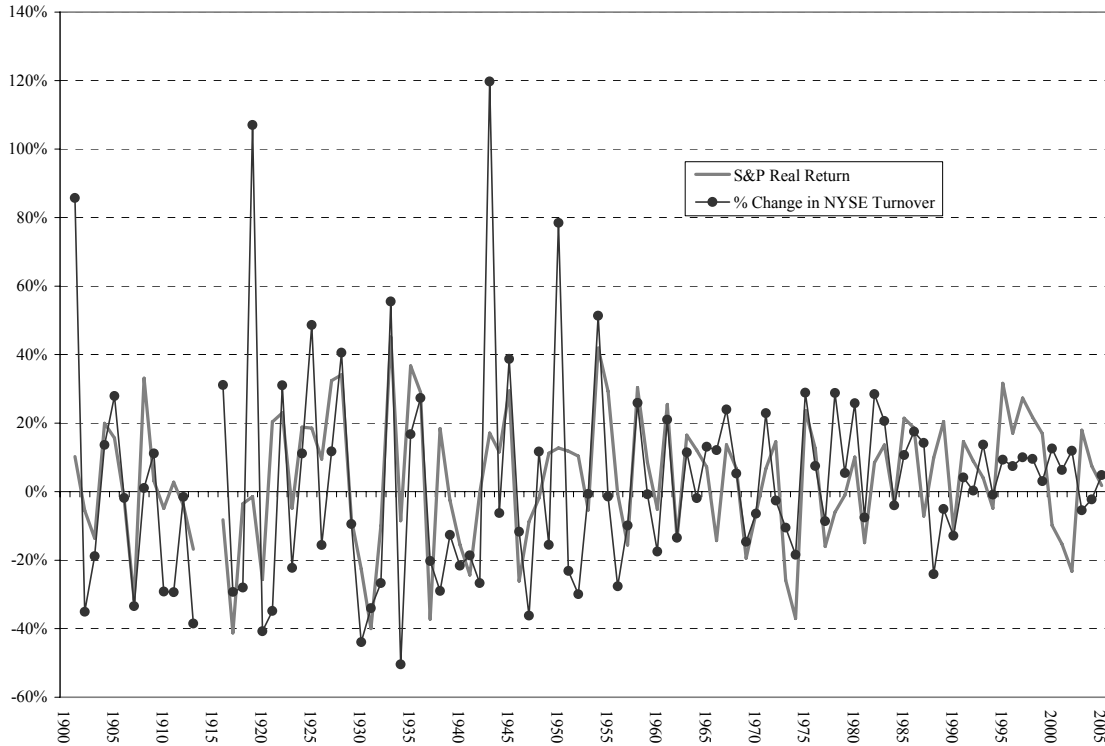
Notes: The underlying data is from the Center for Research in Security Prices (CRSP) database. We use the same sample of internet stocks as Ofek and Richardson (2003). Their sample is obtained from lists of “pure” internet companies published by Morgan Stanley, available on Eli Ofek’s home page at <http://pages.stern.nyu.edu/~eofek/>. For each month, we divide the set of all common stocks listed on CRSP into “internet” and “all other” portfolios, and calculate average monthly turnover and price indices for these two categories. Our turnover and price level indices are equal-weighted, but the results are qualitatively similar using market-capitalization weights.

Figure 2
Turnover in Value and Glamour Stocks, 1986-2005



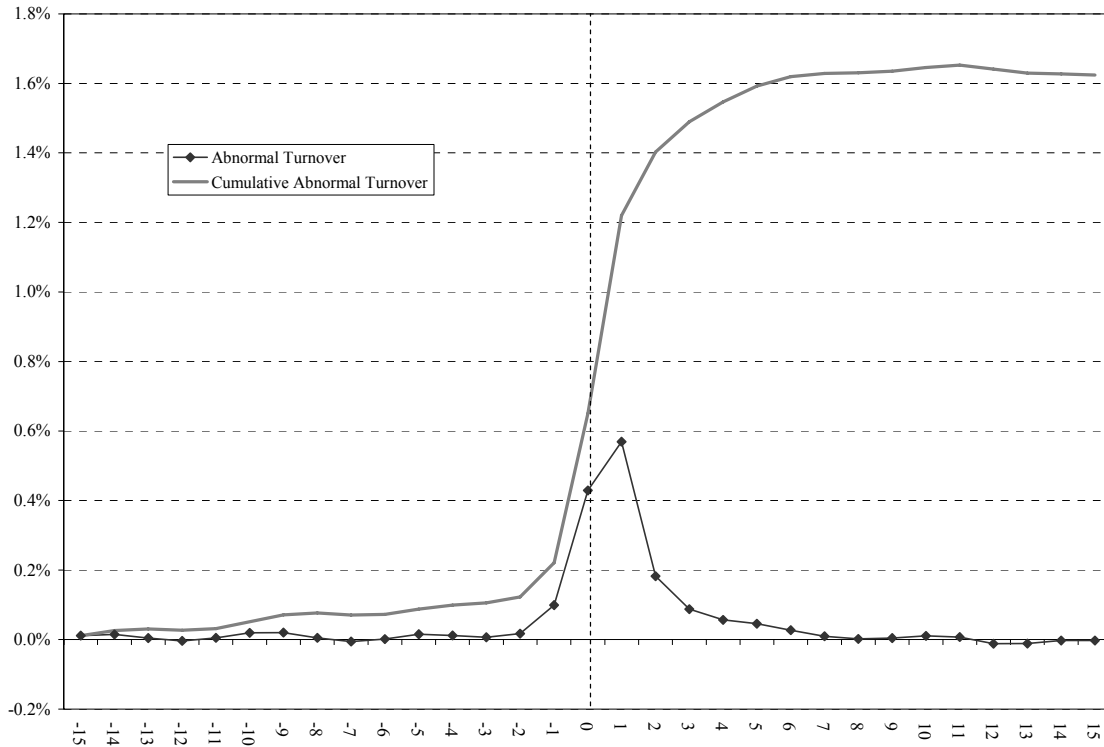
Notes: At any point in time, glamour stocks are the top 200 stocks by market-to-book ratio, out of the 1,000 largest stocks in CRSP. Value stocks are the bottom 200 stocks by market-to-book ratio out of this same universe. For each stock in each month, adjusted turnover is that stock's turnover minus the average turnover of all stocks listed on the same exchange (either NYSE/AMEX or NASDAQ) for that month. This adjustment is done to eliminate differences in reported turnover that are due to the dealer nature of the NASDAQ market. For each month, we then plot average adjusted turnover for the set of glamour and value stocks.

Figure 3
Market-Wide Stock Returns and Changes in Turnover, 1901-2005



Notes: Data on the real (inflation-adjusted) level of the S&P index are from Robert Shiller's webpage, www.econ.yale.edu/~shiller. Data on NYSE turnover are from the NYSE Fact Book, available at www.nyse.com. The figure plots year-to-year percentage changes in the S&P index and year-to-year percentage changes in turnover. We omit the years 1914 and 1915 from the plot since the NYSE was closed for much of the latter half of 1914 due to the outbreak of World War I. The correlation between the two series shown in the plot is 0.49.

Figure 4
Abnormal Turnover Around Earnings Announcements, 1986-2005



Notes: Analysis is based on the universe of the 1,000 largest stocks on CRSP in each quarter from 1986Q1 to 2005Q4. Earnings announcement dates are taken from Compustat. Abnormal daily turnover for any given stock is actual daily turnover in the stock minus average turnover in the stock for the 250 days preceding the event window (days -266 to -16 relative to the earnings announcement date). Cumulative abnormal turnover is then the cumulative sum of abnormal daily turnover over the event window.