Law & Economics Working Papers

Law & Economics Working Papers Archive: 2003-2009

University of Michigan Law School *Year* 2009

Do Individual Investors Affect Share Price Accuracy? Some Preliminary Evidence

Alicia Davis Evans
University of Michigan Law School, alidavis@umich.edu

DO INDIVIDUAL INVESTORS AFFECT SHARE PRICE ACCURACY? SOME PRELIMINARY EVIDENCE

*Alicia J. Davis**
University of Michigan Law School
625 S. State Street
Ann Arbor, MI 48108
(734) 763-2221
alidavis@umich.edu

Draft dated April 8, 2010

* Professor, University of Michigan Law School. For helpful comments on earlier drafts and/or conversations, I thank Omri Ben-Shahar, Laura Beny, Dan Berkowitz, Bernie Black, Nicole Buttermore, Steve Choi, Art Durnev, Jill Fisch, Jian Guo, Jim Hines, Don Langevoort, Jessica Litman, Paul Mahoney, Henry Manne, JJ Prescott, Adam Pritchard, Noah Stoffman, Mark West, Susan Yan, Bernard Yeung, anonymous reviewers and participants at the 2007 American Law and Economics Association Annual Meeting, the 2007 Canadian Law and Economics Association Annual Meeting, the Second Annual Conference on Empirical Legal Studies and the University of Michigan Law and Economics Workshop. For statistical consulting services, I am indebted to Ed Rothman, Brady West, and Heidi Reichert of the University of Michigan's Center for Statistical Consultation and Research. I also thank Paul Michaud and Kyle Schroeder for computer and technical assistance, Michael Ellenbogen, Anthony Rubin, and Caryn Smith for research assistance and Al LaGrone, Karen Rushlow and Matu Zama for administrative assistance. Any errors are my own. The Cook Fund of the University of Michigan Law School provided financial support for this project.

1. Introduction

One goal of securities regulation is protection of the "ordinary" investor (Winter 1988).[1] Some leading scholars, however, argue that ordinary (individual) investors deserve no special protections and, in fact, are "noise traders"[2] that distort share prices (see, for example, Barber, Odean and Zhu 2005).[3] This distortion, if it exists, is troubling because of the important role accurate share prices play in the economy. Stock prices serve as signals for the proper allocation of capital among firms, as investors use stock prices in making investment decisions (Durnev et al. 2003). James Tobin (1982) describes the state of affairs when the stock market directs capital to its highest value use as "functional efficiency." Functional efficiency requires accurate share prices (Durnev et al. 2003).

Because of the importance of share price accuracy, researchers have struggled to understand the factors that affect share price accuracy. The policy implications seem clear: If individuals, as a group, act as noise traders, society might be better served if the direct participation of retail investors in securities markets were restricted.[4] Indeed, Donald Langevoort (2002, 172-73), states:

---

[1] Modern scholars of securities regulation assert that the goal of securities regulation should be the attainment of efficient financial markets to improve the allocation of resources in the economy (see, for example, Goshen and Parchomovsky 2006). Of course, this goal and investor protection do not necessarily conflict. Providing legal protection for ordinary investors is justified if, in doing so, the efficiency of the market functions they perform is enhanced (Winter 1988).

[2] Noise is that which is introduced into stock prices when investors trade, not based on company fundamentals, but on fads, rumors or other types of unreliable information.

[3] Share prices are accurate when they reflect fundamental value by serving as good predictors of future cash flows to shareholders over the life of the firm (see, for example, Fox et al. [2003] for a comprehensive discussion of the related concepts of "share price accuracy" and "share price informedness."). Share prices do this by incorporating information that predicts these future cash flows, rather than reflecting "noise."

[4] Though eliminating individual investors from the capital markets seems politically infeasible, it is not implausible. One leading securities regulation scholar, Stephen Choi (2000), though not arguing that individual investors distort stock prices, has proposed an investor classification scheme (based on informational resources and market knowledge, as displayed on a licensing exam) that would prohibit direct investment in securities markets by unsophisticated investors. In addition, (1) the recent proliferation, and then consolidation, of private trading platforms, which are open only to large institutional investors trading securities issued privately through Rule 144A offerings, and (2) the existence of alternative trading systems (so-called "dark pools"), designed to provide additional liquidity for institutional investors trading in public securities, demonstrate demand for trading venues that exclude individual investors. Though it is perhaps unlikely that a private trading platform or alternative trading systems could replace our current deep, liquid public capital markets, the creation of a market with no or limited individual investor participation is at least possible.

> "…[T]he more emotions and cognitive biases of noise traders adversely affect market prices, the more noise traders can be construed as "bad guys." Good public policy would then be to eradicate these biases if possible, or at least neutralize their social and economic influence…[T]his is the deep concern about where the behavioral literature leads us: if accurate, it invites regulation that privileges the savvy and treats unsophisticated traders as economic undesirables."

This Article provides evidence on the effect of individual investors on share price accuracy. My examination of a new data set of New York Stock Exchange ("NYSE") retail trading statistics shows that retail investor trading levels are significantly correlated with firm-specific stock return variation (also known as R^2), a commonly used measure of share price accuracy. Though there is controversy surrounding the correct interpretation of this metric, the dominant view among R^2 adherents is the greater a stock's firm-specific return variation (that is, the lower the R^2), the more accurate its price. The evidence in this study shows not only that increased levels of retail trading are associated with lower R^2's, but also that there is reason to believe the relationship is a causal one (that is, retail trading causes changes in R^2). Though the implications of these findings depend on the proper interpretation of R^2, the evidence suggests that retail investors have a positive effect on share price accuracy and that restricting the access of large numbers of individual investors to equity markets is not only unwarranted, but also could harm market functioning.

The Article proceeds as follows. Section 2 summarizes the existing literature surrounding noise trading and individual investors in an attempt to provide some context for the debate. Section 3 describes the R^2 methodology, and Section 4 describes the data used and analytical methodologies of this study. Section 4 also presents preliminary results. Section 5 concludes and briefly considers the policy implications of this study's findings.

2. Review of Existing Literature

The effects of irrational or noise traders on stock prices and market efficiency is a central issue in finance scholarship. The traditional belief is that irrational traders cannot affect share prices over the long run. Under this theory, trading based on mistaken beliefs will lead to trading losses against rational, informed investors and wealth reductions that will make it impossible for irrational traders to survive in a competitive marketplace (see, for example, Friedman 1953; Fama 1965). Trades of irrational traders are random and uncorrelated, thus tending to cancel one another out and largely eliminating any price effects from such trading (see, for example, Friedman 1953; Fama 1965). Therefore, as the theory goes, there is little reason to worry about the presence of noise traders.

This traditional view has come under intense theoretical and empirical attack.[5] In a theoretical study, Kogan et al. (2006) conclude that irrational traders can have a persistent effect on stock prices even if they do not "survive" (that is, the value of their trades is infinitesimal in relation to the total value of trades because they have suffered wealth reductions). In addition, a large body of empirical analysis provides evidence that the trades of irrational investors are not random and do not cancel one another out. This evidence suggests that noise traders often act as a herd (see Hirshleifer and Teoh [2003] and Lux [1995] for a review of the literature related to herding behavior). Thus, in the opinion of many in the finance and legal academies, noise trader risk is real, and protecting market efficiency requires creating a climate that can counteract the effects of noise traders (see, for example, Goshen and Parchomovsky 2006; Langevoort 2002).[6]

Most prior research shows that individual investors are more likely to make irrational or imprudent investment decisions than institutional investors are (Jackson 2003). Barber, Odean and Zhu (2005) find that small trades (serving as a proxy for trades by individual investors) are correlated, buying by retail investors pushes prices too high (above their fundamental values), and selling by retail investors

---

[5] As noted by Kogan et al. (2006, 195-96), De Long et al. (1990) argue, on the basis of a partial equilibrium model, that irrational traders hold portfolios with high growth and can potentially outgrow rational traders and thus survive. Conversely, Sandroni (2000) and Blume and Easley (2001), using general equilibrium models, conclude "irrational traders do not survive in the long run."

[6] However, see Section 5 for a discussion of the work of scholars that suggests that noise traders can enhance market functioning.

pushes prices too low (below their fundamental values). They conclude that individual investors (whom they term noise traders) can move equity markets. Similarly, Hvidkjaer (2006), using trade size as a proxy for the trades of individual investors, conducts a study on individual investor trading patterns and concludes that there is a systematic component to retail trading and that such trading behavior can lead to or protract periods where a stock is over- or undervalued. Kumar and Lee (2006) also find that retail investor trades are systematically correlated, that this concerted action can affect stock returns and that this "retail sentiment" does not appear to be an outgrowth of a reaction to factors related to fundamental value.

But evidence from international markets calls into question the idea that retail investors are the noise traders in markets. Jackson (2003), after analyzing a unique dataset of 41.9 million retail investor trades over an 11-year period on the Australian Stock Exchange,[7] finds that, though individuals invest in a systematic fashion, it would not be appropriate to characterize their trading behavior as "irrational." Indeed, the trades of individuals investing through full-service brokerage firms positively predict future market returns.[8] Jackson states that one potential reason for this result could be individuals' possession of valuable private information.

In addition, Choe, Kho and Stulz (2001), based on a study of two years (1996-1998)[9] of Korea Stock Exchange trading data, find that domestic retail investors[10] possess a short-lived informational advantage over both foreign investors and domestic institutional investors. An event study on trading behavior of different classes of investors around days on which stock prices have a 5% or more, in

[7] According to Jackson, as of 2001, the U.S. market is 37 times as large as the Australian market (as measured by total market capitalization). Also, in the U.S. retail investors own approximately 42% of the stock traded on U.S. markets. In Australia, the percentage of individual investor ownership is 24%. The U.S. and Australian markets are not strictly comparable, but Australia is one of the few markets in the world with sizable retail investor participation. Thus, Jackson's study results should be interesting to those concerned about the effect of individuals on market functioning.

[8] Jackson examines trades from individuals that invest through 47 full-service brokerage firms and 9 Internet brokers. Jackson reports that the trades of the full-service brokerage clients drive the study's results.

[9] The Asian crisis of the late 1990's and its market effects during the study period may affect the researchers' findings and make the results ungeneralizable.

[10] Choe, Kho and Stulz (2001) report that, at the time of the study, domestic retail investors were the most active traders on the Korea Stock Exchange, with their sales representing 77.4% of the gross value of stock sales in 1998. This is a much higher proportion of retail investor trading than in the United States.

absolute value, abnormal return reveals that domestic individual traders have a higher proportion of buy trades before the event than after and a lower proportion of sell trades before the event than after. No other investor class studied exhibits this pattern. These results suggest that individual investors in the aggregate are capable of predicting future corporate events.[11]

This Article makes two new contributions to the literature. First, it provides information to assess the noise trader claim against retail investors. One of the benefits of this study is its use of direct New York Stock Exchange retail trading data and not individual investor trading proxies such as small share size[12] or odd-lot trading.[13] Those proxies suffer from some important limitations.[14] First, using small share size as a proxy for trades by individual investors potentially can distort results. Institutions often break up their trades into smaller batches to hide their intentions from other market participants or for other, liquidity-based reasons (Hvidkjaer 2006). Thus, trades that appear to be made by individuals, given the small size, could actually be a portion of a trade made by an institutional investor.[15] In addition, odd-lot trading data have a strong potential for underinclusiveness with respect to retail trading. Though any investor may trade in odd lots, a common belief is that individual investors of lower wealth are more apt to do so. Using odd-lot trading as a proxy for the individual investor trading is problematic, however, because many individual investors trade in round lots; indeed, there is evidence that they prefer

[11] The database on which Choe, Kho and Stulz (2001) rely does not distinguish between public market individual investors and company insiders. Insiders, of course, may avail themselves of non-public information before trading, which, of course, would make it appear as though they can predict corporate events.

[12] Barber, Odean and Zhu (2005) recognize the limitations of small share size as a proxy for individual investor trading and do some limited testing of their data against actual brokerage firm data to gain comfort in the representativeness of their data sets.

[13] Odd-lot trading is trading a number of shares other than that which is required for a round lot (100 shares). Wu (1972), in a study of individual investor trading behavior that used odd-lot trading data as a proxy for the trades of such investors, concludes that odd-lot trading has no effect on share prices. For an example of a recent work using odd-lot trading data as a proxy for retail investor trading, see Brusa, Liu and Schulman (2005).

[14] Kaniel, Saar and Titman (2008), in a study of retail investor behavior, use a proprietary data set of NYSE trades provided to the researchers directly by the NYSE. They find that retail investor trades tend to follow certain patterns, but that there is no strong evidence that the trades of individuals are correlated. Though their findings are relevant to the ongoing debate, they do not provide evidence that directly addresses the question of whether individual investors affect share price accuracy.

[15] Indeed, the average size of a market trade today is 260 shares, down from 1,400 shares a decade ago (Patterson and Lucchetti 2008).

to do so.[16] Use of odd-lot data loses the impact of round-lot trading in study results. This study's second contribution to the literature is its examination of a relationship that has not yet been studied -- the relationship between retail trading activity and R^2.

3. R^2

In this study, I use "R^2" as a measure of share price accuracy. "R^2" is the R^2 statistic obtained by regressing the individual returns of a firm's stock on the returns of the market as a whole and the firm's industry group (excluding the firm in question). In statistics, R^2 tells how much of the variation (expressed as a percentage) observed in a dependent variable (in this case, a firm's individual stock return) is explained[17] by the independent or explanatory variables (in this case, the market return and the industry return). R^2 takes the value of 0.0 – 1.0, with an R^2 of 0.0 signifying that none of the variation in the dependent variable is explained by its relationship with the independent variables. Conversely, an R^2 of 1.0 means that 100% of the variation in the dependent variable is explained by its relationship with the independent variables.

In this study, a high R^2 means that much of the variation in an individual firm's stock returns can be explained by the market return and the industry return. In other words, the firm's stock price is influenced primarily by movements in the market as a whole and stocks in the firm's industry group. Conversely, a low R^2 means that a firm's stock price movements bear little relation to movements in market prices or stock prices of its industry peers. There are two potential explanations for this occurrence: 1) Low R^2 means that the firm's stock price is more "informationally efficient" because it incorporates firm-specific (rather than market or industry) information and is therefore more "accurate" or 2) Low R^2 means the firm's stock price reflects significant non-market or industry information, but such

[16] See note 56 for further discussion.
[17] The word "explained" as used in this context does not suggest that the independent variables *cause* changes in the dependent variable.

information is noise, rather than information related to a company's fundamentals (Roll 1988, Durnev et al. 2003).[18]

In current finance scholarship, among R^2 metric adherents,[19] the former view prevails, but is controversial. Resolving the R^2 controversy is important for interpreting the finding of this study and others that rely on the R^2 methodology. What follows is a review of R^2-related studies. A number of prior studies suggest that R^2 is consistent with greater share price accuracy, but several researchers have found evidence that leads them to reach the opposite conclusion. Recent work by Dasgupta, Gan and Gao (2010) provides a path for reconciling the competing interpretations of R^2, and, as I describe in Section 5, the instant study provides evidence consistent with their theory.

Available evidence that a low R^2 is a metric of informational efficiency and share price accuracy is strong. Durnev et al. (2003) provide the most direct evidence on this question as they compare firm-specific stock return variation (R^2) and accounting-based measures of stock price informativeness. They define stock price informativeness as the measure of how much information stock prices contain about future earnings, estimated from a regression of then-current stock returns on current and future accounting earnings. Durnev et al. (2003) find that firm-specific variability (a lower R^2) is positively correlated with their measures of stock price informativeness and conclude that low R^2 is indeed a sign of share price accuracy and not noise impounded in share prices.

In addition, Durnev, Morck, and Yeung (2001) find that firms operating in U.S. industries with lower R^2's use more external financing. The authors suggest that this relationship is evidence that low R^2 is associated with stock prices that more closely track firm fundamentals. In another study, Durnev, Morck, and Yeung (2004) find a strong correlation between firm-specific return variation and

---

[18] Of course, a low R^2 also could mean that a firm's stock price moves largely independently of the market and its industry group and that the information compounded in its stock price is a combination of fundamental information and noise. This explanation is quite plausible, but, as described in this section, researchers have provided evidence on whether one or the other explanation is more likely, rather than a little of both.

[19] One may argue that R^2 is very similar to beta. A high beta indicates that a stock's returns closely track the returns of the overall market. Under asset pricing theory, both high beta and low beta stocks operate in an environment that is assumed to be fully informed. Thus, some may be skeptical of the claim that the R^2 is a measure of share price accuracy. However, I proceed under the assumption that the level of R^2 reflects informational efficiency because this interpretation has been used extensively in recent finance scholarship.

economically efficient corporate investment. They suggest that capital investment should be more efficient when stock prices are more informative because accurate prices give signals to both management and financial market participants about the quality of management's investment decisions. Presumably, managers may use this signal to change course when necessary, and investors, as Durnev, Morck, and Yeung (2004) suggest, may use this signal to intervene as necessary in the face of poor management decisions. Similarly, Chen, Goldstein and Jiang (2007) present evidence that fluctuations in stock prices affect the capital investment decisions of firms with low R^2's more than those with high R^2's. The authors of the study conclude that this result serves as evidence of managers learning valuable information about company fundamentals from changes in stock prices and incorporating such new knowledge in their investment decisions.

The conclusions the researchers draw in these studies of informational efficiency at the firm and industry level are consistent with evidence of R^2 at the country level. A number of studies show correlations between better functioning equity markets and greater firm-specific return variation (lower R^2's). For example, Morck, Yeung and Yu (2000) calculate, inter alia, the average R^2's of the firms in each of 40 different countries. The five countries with firms having the highest average levels of firm-specific variation (lowest R^2's) are, in order, the United States, Ireland, Canada, the UK and Australia. The five countries with firms having the lowest average levels of firm-specific variation (highest R^2's) are, in order, Poland, China, Malaysia, Taiwan and Turkey. Overall, and with few exceptions, firms in low-income economies (measured by per capita GDP) have, on average, the highest R^2's. This evidence is consistent with the intuition that firms in more well developed economies have more accurate share prices. Similarly, researchers find that lower average firm R^2's are associated with more efficient capital allocation in a country (Wurgler 2000) and less country-level opaqueness[20] (Jin and Myers 2006).

The R^2 methodology also has found adherents among the ranks of legal scholars. For example, Fox, Morck, Yeung and Durnev (2003) conclude that enhanced mandatory disclosure rules adopted in the

[20] Jin and Myers (2006, 281) define opaqueness as a "lack of information that would enable investors to observe operating cash flow and income and determine firm value."

United States in December 1980 made share prices more accurate (as evidenced by a decrease in average R^2 across firms). Beny (2005) employs the country R^2 statistics of Morck, Yeung and Yu (2000) as a measure of share price informativeness and concludes that stronger formal insider trading laws in a country are associated with more informative share prices (that is, lower average R^2) for firms within that country.

Though a number of leading finance and legal studies use R^2 as a metric of stock price informational efficiency, Hou, Peng and Xiong (2006) question the informational-efficiency interpretation of R^2 and contend instead that R^2 is a measure of price *inefficiency*. They base this view on the results of independent empirical analysis and also point to the following studies as evidence consistent with their interpretation of R^2.

Chan and Hameed (2006) compare stock price synchronicity and research analyst activity in emerging markets and find that greater coverage by research analysts is associated with more stock price synchronicity in the market (lower firm-specific information in prices or higher R^2's).[21] Based on these results, Chan and Hameed (2006) conclude that the conventional wisdom that research analysts produce firm-specific information is incorrect and that what analysts actually produce is market-wide information. Hou, Peng and Xiong (2006), however, point to this result as evidence that the R^2 metric is not a measure of informational efficiency because one expects analysts to produce firm-specific information.

The work of Veldkamp (2006) may help bridge the gap between these two opposing conclusions. Veldkamp (2006, 823) sets forth a model in which "investors purchase information that generates comovement." Consistent with this model, she argues, is the finding by Hameed, Morck and Yeung (2005) that firms (after controlling for size) with more analyst coverage tend to have fundamentals that predict other firms' fundamentals. Thus, the information provided by research analysts can produce comovement. Veldkamp's analysis suggests that finding that higher average R^2 in a country is associated

[21] This result is consistent with that of a similar study of firms in the United States (see Piotroski and Roulstone 2004).

with greater research coverage does not mean that it is implausible for low R^2 to be consistent with more accurate share prices.

There is additional evidence, however, that calls the prevailing interpretation of R^2 into question. Hou, Peng and Xiong (2006) cite a study performed by Ashbaugh-Skaife, Gassen and LaFond (2006), which finds that (1) higher, not lower, R^2's are associated with more informative prices in the United States and Germany and (2) no statistically significant relationship exists between R^2 and measures of stock price informativeness in the U.K., Australia, France or Japan. Ashbaugh-Skaife, Gassen and LaFond conclude that there is no consistent relationship between R^2 and stock price informativeness in international markets.

Teoh, Yang, and Zhang (2006) find that firms with low R^2's are more likely to have accounting-based return anomalies, poor earnings quality and weak fundamentals. Thus, rather than a metric of share price accuracy, these researchers assert that a low R^2 is actually an indicator of the level of *uncertainty* faced by investors. Hou, Peng and Xiong (2006) find evidence that stocks with lower R^2's exhibit what the researchers term "overreaction-driven price momentum" and more long run price-reversals.[22]

Finally, legal scholar Ferrell (2003) reports the change of R^2 in over-the-counter stocks (OTC) after passage of the 1964 Securities Acts Amendments in the United States. The 1964 Amendments extended mandatory disclosure requirements to over-the-counter stocks. Before the 1964 amendments, these requirements applied only to exchange-listed stocks. Ferrell finds that before OTC-mandated disclosure, the R^2's of OTC stocks were lower, on average, than those of listed stocks. Ferrell states that it is "highly implausible" that the OTC market was more informationally efficient than the listed market before the 1964 amendments.

Ferrell likely finds this outcome implausible because of the limited disclosure from OTC market firms before the amendments. Ferrell reports that, in 1963, the Securities and Exchange Commission

[22] The researchers consider price momentum in a stock an outcome of investor overreaction. Price momentum, a sign of informational inefficiency, is defined in their study as the presence of a phenomenon by which an investor could buy the "winning" stocks and short sell the "losing" stocks from a prior six-month period and generate economically and statistically significant trading profits over the next one to six months.

(SEC) completed a report on the state of securities regulation. In its report, the SEC found that, of a random sample comprising approximately 20% of all OTC companies, 25% of the firms did not supply any financial data to their shareholders at all, and 23% did not certify their financial statements. Of those firms that did provide financial data to their stockholders, 44% did not categorize their inventories, and 33% failed to provide explanatory notes on significant financial matters, including depreciation methods, long-term contractual obligations and contingent liabilities, all of which firms with listed stocks were required to disclose.

As demonstrated above, the prevailing view in the literature among R^2 adherents is that R^2 is a measure of price informativeness, but that view is controversial. Lee and Liu (2007) attempt to reconcile the two competing interpretations of R^2. They hypothesize that the relationship between R^2 and share price informativeness is not monotonic, but rather U-shaped. They thus argue that when there is more firm-specific return variation in stocks that operate in good information environments,[23] lower R^2's are a sign of more price informedness. Conversely, for firms that operate in poor information environments, higher R^2's are a sign of more accurate prices.[24] The idea is that if a firm operates in a good environment for information, then having more of the firm-specific variety will enhance share price accuracy.[25]

Dasgupta, Gan and Gao (2010) also set forth a theory that is highly persuasive in reconciling the competing views on R^2. The researchers argue that the R^2 metric must be put into context vis-à-vis a firm's transparency before the period over which R^2 is calculated. For example, consider, the researchers urge, the extreme case of Firm ABC that is "completely transparent" as of December 31, 2006 (that is, there is no firm-specific information about Firm ABC that is unknown to public market investors as of

[23] Lee and Liu define a good information environment for a firm as one characterized by (1) higher institutional ownership, (2) longer time in existence, (3) lower research analyst forecast dispersion, (4) lower research analyst forecast error, (5) higher liquidity (greater ease of selling without affecting price) and (6) a lower probability that a market maker in a stock will trade with an informed trader (because most firm-specific information is already incorporated into the price). The last characteristic is derived from a market microstructure model first set forth by Easley, Hvidkjaer, O'Hara (2005). Interpreting lower values of this metric as a sign of greater price informativeness is controversial (see Lee and Liu 2007).

[24] The researchers rely on the information environment rather than the metric of informativeness used by Durnev et al. (2003) (that is, how well the price predicts future earnings) because of data availability and tractability given their research design.

[25] See Teoh, Yang, and Zhang (2006) for evidence that calls into question certain elements of Lee and Liu's (2007) theory of the U-shaped relationship between price informativeness and idiosyncratic volatility.

that date). In this case, investors would have incorporated perfectly firm-specific data into the then-current price. If researchers were to regress the returns of Firm ABC on the market's return, say, for example, from January 1 – December 31, 2007, they likely would find, according to the account of Dasgupta, Gan and Gao (2010), that the regression would yield an R^2 of 1.0. The authors argue that as firm-specific events unfold during 2007, the firm's stock price would not move after these events because the market already would have anticipated the occurrence of such events (that is, there is no "surprise").[26] The only area of uncertainty with respect to the appropriate price for Firm ABC's stock is the effect of market-wide events on Firm ABC, which would explain the perfect comovement with the market return. Conversely, a firm that is "completely opaque" on December 31, 2006, likely would have a low R^2 for the January 1 – December 31, 2007 period. Thus, the researchers posit that firms that operate in better information environments are likely to have higher return synchronicity with the market.[27]

4. Data Sources, Sample, Methodology and Results of Analysis

4.1 Data and Sample

The period of this study is April 1, 2005 – August 31, 2006. I obtained data on firm-level, industry-level and market returns, as well as share prices, shares outstanding, total volume and firm industry group from the Center for Research in Security Prices (CRSP) database. I acquired firm-level accounting data from the merged CRSP-Compustat database.[28] I obtained data on firm-level retail trading activity, including total shares purchased and sold by retail investors on the NYSE each day[29] for a particular stock[30] from the NYSE ReTracEOD Summary.[31] Institutional ownership data, based on the

---

[26] This analysis assumes, of course, that the market would be able to incorporate the information into the price accurately.
[27] Dasgupta, Gan and Gao (2010) also find empirical support for their theoretical assertions. They find that older firms have higher R^2's, and that stock return synchronicity becomes significantly higher following equity issuance-related disclosures.
[28] See Section 4.3 for a description of the accounting data used in this study.
[29] Retail trading data for April 21, 2006 are unavailable and, thus, are not a part of the sample. However, because there is no reason to believe that retail trading behavior differed significantly on that date from other dates in the sample, this omission should not bias the overall conclusions of this Article.
[30] Though this figure does not represent all trading by retail investors in NYSE-listed stocks (only that executed on the NYSE), the study data should capture the overwhelming majority of the retail trading activity in NYSE-listed stocks. Researchers estimate that between 75% and 85% of trading volume of NYSE-listed stocks is executed on

quarter ended March 31, 2005, are from the Thomson Financial institutional holdings database. The First Call database is the source of information on research coverage and research activity during the study period, and I derived news coverage information from Dow Jones News Service articles. Finally, I obtained data on industry SIC codes, insider ownership and 5% holder ownership, as of March 31, 2005, from Thomson Financial's Compact D SEC Disclosure.

To construct a sample, I began with every NYSE-listed common stock[32] in the CRSP database during the study period.[33] I then eliminated from the sample any company that lacked data for the variables used in the study and all firms in industry groups with fewer than three members.[34] My final sample for this study consists of 1,129 different stocks. For all analyses, I report results derived from the overall sample, as well as results derived after splitting the sample into two approximately equal groups of 565 and 564 different stocks based on size (as determined by average market capitalization during the study period). I refer to these groups as "Top Half" (the larger firms) and "Bottom Half" (the smaller firms).

---

the NYSE (see, for example, Goldstein et al. 2006). There is no reason to believe that retail trades executed on the NYSE do not represent a similar proportion of overall retail trading activity in NYSE-listed shares.

[31] The NYSE generates ReTrac figures from information accompanying orders. Every order executed on the NYSE must have an account-type designation (see ReTracEOD Data Discussion Board 2006; New York Stock Exchange 2004, 29). ReTrac EOD files track retail investor trades (defined as those made by accounts with the designation "I" (non-program trading, individual investor, as defined in NYSE Rule 80A)). NYSE Rule 80A offers the following definition: "'Account of an individual investor' means an account covered by Section 11(a)(1)(E) of the Securities Exchange Act of 1934." Exchange Act Section 11(a)(1)(E) covers the following accounts: "… the account of a natural person, the estate of a natural person, or a trust (other than an investment company) created by a natural person for himself or another natural person." It is possible for brokers to execute individual investor trades along with institutional investor orders (and without the "I" designation), which could result in information from such retail trades not being included in the ReTrac data. However, this occurrence generally is believed to be rare.

[32] Stocks include those classified by CRSP as "ordinary common shares" of share codes 10 (companies that have not been further defined), 11 (companies that need no further definition), 12 (companies incorporated outside the U.S.) and 18 (REITs).

[33] Firms that fail to trade on any day during the study period are excluded. When a firm does not trade on a particular day, CRSP gives its daily return a value of "0." Including these firms in the sample would distort the R^2 calculation because the "0" value is not a reflection of investors' collective decision to keep the stock at the same price after a day of trading, but rather the result of no trading activity at all.

[34] I cannot calculate R^2 for firms operating in industries with fewer than three members because I cannot construct an "industry group" of two or more firms for use in the calculation.

4.2 Firm-Specific Stock Return Variation

As a measure of share price informedness or accuracy, I use firm-specific stock return variation. Following the model of Durnev et al. (2003) and others, I obtain firm-specific stock return variation (R^2) by use of the following regression:

$$r_{i,d,t} = \alpha_{i,t} + \beta_{i,t}\, r_{m,d,t} + \gamma_{i,t}\, r_{j,d,t} + \varepsilon_{i,d,t} \tag{1}$$

of firm *i*'s total returns $r_{i,d,t}$ on market return $r_{m,d,t}$ and a broad industry return $r_{j,d,t}$, which includes the market value-weighted average return of all firms in industry *j* (defined as all firms in the same two-digit SIC code), excluding the firm in question.[35] Returns are measured across *d* daily periods during the study period *t* (April 1, 2005 – August 31, 2006). If the prevailing view in the literature among R^2 adherents is correct, the lower the value of R^2 generated from the above regression, the more firm-specific information there is incorporated into a firm's stock price and the more accurate the price.

4.3 Empirical Methodology and Results of Analysis

My objective is to examine the relationship between firm-specific stock return variation and retail trading activity. Consistent with the practice in the R^2 literature, I use the logistic transformation[36] of R^2, New R^2, as my dependent variable.[37] I obtain New R^2 by using the following formula:

$$New\ R^2 = ln(R^2_i/1 - R^2_i) \tag{2}$$

In the regressions that follow, my independent variable of interest is retail trading. (See Table 1 for descriptive statistics for this study.) I define retail trading activity as the proportion of the trading in a firm's common stock that is executed by retail investors. I calculate two measures of retail trading activity in this study: 1) the ratio of the number of shares of a firm's stock bought by retail investors each trading day in the study period to the total number of a firm's shares traded each day in the study period,

---

[35] Consistent with Durnev et al. (2003), I exclude the firm in question to avoid spurious correlations between firm returns and industry returns for companies in industries with only a few firms.

[36] This is a common econometric remedy (see, for example, Morck, Yeung, and Yu 2000). The transformed variable is a continuous variable that is more normally distributed than R^2, which has values between 0 and 1 (Ashbaugh-Skaife, Gassen and LaFond 2006).

[37] In unreported results, I run the regressions described in Section 4.3 below using R^2, instead of New R^2. My estimates in this alternative regression are less precise, but qualitatively my results are similar to those described in Section 4.3.

averaged over the study period ("buy-side retail trading") and 2) the ratio of the number of shares of a firm's stock sold by retail investors each trading day in the study period to the total number of a firm's shares traded each day in the study period, averaged over the study period ("sell-side retail trading").

I control for variables that may affect share price accuracy. I control for level of "INSTITUTIONAL OWNERSHIP" (defined as the proportion of a firm's stock held by institutions as of March 31, 2005)[38] because there is reason to believe that firms with a largely institutional shareholder base may be more likely, in response to shareholder demand, to provide specific earnings guidance and make voluntary disclosures about their business prospects that can help the market more accurately establish prices for the firms' shares. My control variables also include "SIZE" (measured as a firm's average market capitalization (closing share price x number of shares outstanding) during the study period) and "VOLUME" (average daily volume of total shares traded in a firm's stock during the study period) because larger, more liquid firms are more likely to have a large investor following and generate more interest and, potentially, private information. I account for the effects of research analysts who disseminate firm-specific information into the marketplace with two variables: 1) the number of different analysts that cover the firm (as evidenced by the publication of earnings estimates) during the study period ("RESEARCH COVERAGE") and 2) the total number of earnings per share estimates released by analysts for a firm during the study period ("RESEARCH ACTIVITY"). Because controllers or other insiders may possess superior private information about a firm's prospects, but also, conversely, may make firms less transparent (see, for example, Pizarro et al. 2007), I control for the proportion of a firm's stock held by insiders (for example, directors and officers) ("INSIDER OWNERSHIP") and for the proportion of a firm's stock held by individuals or institutions with 5% or greater stock ownership in the

[38] One should note that because the institutional ownership variable is based on Form 13-F data (all institutions with $100 million or more in securities under discretionary management are required to report their holdings to the SEC each quarter), the variable only represents stock owned by large institutions (that is, those with $100 million or more in assets under management) and does not account for shares held by small institutions. One also should note that because of duplicative reporting by institutions on the required Form 13-F's some firms in the study sample have institutional ownership percentage values that, as calculated, exceed 100% (Other researchers have found that such instances of duplicative reporting are generally rare (see, for example, Thakor, Nielsen and Gulley 2005) and, thus, the figures, though anomalous, should not bias this study's results significantly. In the instant study, 65 of the 1,129 stocks in the sample (5.8%) have institutional ownership percentages that, as calculated, exceed 100%).

company ("5% OWNER OWNERSHIP").[39] Diversified conglomerates may be more difficult for investors to understand and value or may track the market more closely because they operate in more segments of the economy. I, therefore, control for firm-level diversity ("DIVERSITY"), measured by the number of four-digit SIC codes in which the firm operates.[40] In addition, I control for firm news coverage ("NEWS COVERAGE") during the study period[41] because media attention affects the amount of firm-specific information in the marketplace. News coverage also may affect the trading behavior of retail investors, as Barber and Odean (2008) find that individual investors are attracted to stocks that capture their attention for a number of reasons, including by being featured in news stories.

I also employ a number of additional controls suggested by the work of Baker and Wurgler (2006) that are related to share price accuracy. Baker and Wurgler argue that stock mispricings result from "both an uninformed demand shock and a limit on arbitrage." Investor sentiment, they argue, may vary across firms and affect prices in the following manner. Using one possible definition of investor sentiment as the "propensity to speculate," Baker and Wurgler suggest that investor sentiment may drive

---

[39] A more apt variable would be the proportion of trades represented by insiders or 5% holders, rather than ownership by such investors, but this trading information is not currently available in a comparable format to that of the retail trading data I use in this study.

[40] In unreported results, I use, as an alternative, four dummy variables in my regression representing the four-categories of diversification suggested by Varadarajan and Ramanujam (1987). Under this formulation, I calculate a firm's level of "broad spectrum diversity (BSD)," measured as the number of two-digit SIC code industries in which a firm operates and a firm's level of "mean narrow spectrum diversity (MNSD)," defined as the number of four-digit SIC code industries in which a firm operates. Firms with BSD levels below the mean of my sample population are characterized as "low" BSD firms, and firms with BSD levels above the mean of my sample population are characterized as "high" BSD firms. Similarly, firms with MNSD levels below the mean of my sample population are characterized as "low" MNSD firms, and firms with MNSD levels above the mean of my sample population are characterized as "high" MNSD firms. Firms that have both low BSD and low MNSD are "Category 1" firms or those with "very low diversity." Firms that have both high BSD and high MNSD are "Category 4" firms or those with "very high diversity." Firms with low BSD, but high MNSD are "Category 2" firms or "related-diversified" firms. Firms with high BSD, but low MNSD are "Category 3" firms or "unrelated-diversified" firms. My results remain qualitatively unchanged when I employ these dummy variables instead of the one based simply on the number of four-digit SIC codes.

[41] Consistent with prior studies, I calculate level of news coverage by hand counting the number of days during the study period on which a firm is featured prominently in a Dow Jones News Service story. "Featured prominently" means being mentioned by name either in the headline or lead paragraph. I use number of days of coverage, rather than the number of individual news stories, to avoid the possibility of counting multiple, essentially identical stories appearing in the Dow Jones News Service on the same day. Counting duplicate news stories could provide a distorted view of the amount of information disseminated to the public marketplace. Note, however, in a separate regression, I use the raw number of news stories over the study period as the "news coverage" independent variable. In unreported results, I find the outcome does not change qualitatively from the results in the original regression as reported in this Section 4.3.

the demand for speculative investments. They also argue that what makes a firm's stock particularly vulnerable to investors' propensity to speculate lies in large part in the subjectivity of the firm's valuation. Firms that are young and unprofitable and that have extreme growth prospects allow unsophisticated investors to defend "with equal plausibility" a wide range of valuations that are consistent with the investors' general market sentiment (that is, either general pessimism or optimism). This form of speculation is more difficult to do with firms with a long, established earnings history, tangible assets and stable dividends.

Similarly, firms with characteristics that make arbitrage (to offset any noise trading or speculative tendencies) difficult are also more subject to mispricing. Drawing on prior research that shows that arbitrage is particularly costly and risky for stocks of young, small, unprofitable, extreme growth or distressed firms, Baker and Wurgler posit that such firms are more likely to be mispriced.[42] The stocks that are the hardest to value are also the most difficult to arbitrage.

I therefore adopt the following additional controls: "AGE" (firm age, measured, to the nearest month, as the number of months the firm has appeared in CRSP), "VOLATILITY" (stock volatility, measured as the standard deviation of monthly stock returns over my 17-month study period),[43] two variables related to profitability -- "RETURN ON EQUITY"[44] and a dummy variable for whether a firm is profitable (that is, has positive earnings), two variables related to dividend payments -- "DIVIDENDS-

[42] As Baker and Wurgler (2006, 1649-50, citations omitted) explain, "First, [the stocks'] high idiosyncratic risk makes relative-value arbitrage especially risky. Moreover, such stocks tend to be more costly to trade and particularly expensive, sometimes impossible, to sell short. Further, their lower liquidity exposes would-be arbitrageurs to predatory attacks."

[43] One could argue that a useful measure for comparison purposes would be relative standard deviation of returns, calculated as standard deviation divided by the mean, rather than the raw standard deviation as used by Baker and Wurgler. Relative standard deviation is used to compare the variability of data when the means of the data (here, firms' average stock returns) are significantly different across the sample. I run the regression analysis described in this Section 4.3 using both standard deviation and relative standard deviation and in unreported results find that though standard deviation is a significant variable in the regression, relative standard deviation is not. However, the relationship between retail trading (the variable of interest) and R^2 is qualitatively identical in both formulations.

[44] Return on equity is defined as earnings/book equity. Earnings (E) is income before extraordinary items (Compustat Item 18) plus income statement deferred taxes (Compustat Item 50), minus preferred dividends (Compustat Item 19). Book equity (BE) is stockholders' equity (Compustat Item 60) plus balance sheet deferred taxes (Compustat Item 35). All references to Compustat Item numbers in this note and the ones to follow, unless otherwise noted, are for the year 2004 (that is, the year immediately before the beginning of the study period).

TO-EQUITY"[45] and a dummy variable for whether a firm pays dividends (that is, has positive dividends per share), two variables related to asset tangibility -- the ratio of "TANGIBLE ASSETS TO TOTAL ASSETS" and the ratio of "RESEARCH AND DEVELOPMENT EXPENSES TO TOTAL ASSETS,"[46] and three variables to proxy for characteristics indicating high growth opportunities and/or distress -- "BOOK-TO-MARKET EQUITY,"[47] "LEVEL OF EXTERNAL FINANCE,"[48] and "SALES GROWTH."[49] Table 2 contains pairwise correlation coefficients for the independent variables used in this analysis.

After performing a regression analysis using firm-specific return variation (R^2) as the dependent variable, and all of the above explanatory variables, I run several types of regression diagnostics. A check of the normality of residuals and the existence of a linear relationship between the dependent variable and the explanatory variables reveals that a number of the independent variables are problematic. Thus, to correct for this deficiency where warranted, I express the values of certain variables in natural logarithms. The variables requiring this treatment and for which I provide the requisite transformation[50] include buy-side retail trading, sell-side retail trading, market capitalization, trading volume, research activity, insider

---

[45] Dividends-to-equity is defined as dividends/book equity. Dividends are the dividends per share at the ex date (Compustat Item 26) times Compustat shares outstanding (Compustat Item 25) divided by book equity, as defined in note 44 above.

[46] Tangible assets to total assets is defined as property, plant and equipment (Compustat Item 7) divided by total assets (Compustat Item 6). Research and development expenses to total assets is defined as R&D expense (Compustat Item 46) divided by total assets (Compustat Item 6). Consistent with Baker and Wurgler (2006), missing values of R&D expense are set to zero. As a robustness check, I run the regression analysis described in this Section 4.3 first, including all firms in the sample and then, excluding firms with missing R&D values. In unreported results, I find that the outcomes are qualitatively identical.

[47] Book to market equity is book equity (defined in note 44 above) divided by a firm's average market capitalization during the study period.

[48] External finance is defined as the change in assets (Compustat Item 6) from 2003 to 2004 minus the change in retained earnings (Compustat Item 36) over the same period divided by total assets (Compustat Item 6).

[49] Sales growth is the change in net sales (Compustat Item 12) from 2003 to 2004 divided by 2003 net sales. In addition, the firms in the sample are divided into deciles by sales growth, with firms with the highest level of sales growth considered "extreme growth" firms. In addition, firms in the top three deciles are considered "high growth" firms, and firms in the bottom three deciles are considered "low growth firms."

[50] The variables for return on equity, sales growth, and external finance show some evidence that they require this treatment, as well, but the conclusion is not clear. I, therefore, leave these variables in the regression in their unaltered state. I run a regression analysis using transformed variables for return on equity, sales growth and external finance, and my overall results remain qualitatively unchanged from those I report in this Section 4.3. I, however, lose almost 500 observations. I lose such a large portion of my sample following this transformation because a significant number of the firms in my sample have negative values for these variables, for which a logarithmic transformation is impossible.

ownership, dividends-to-equity, R&D-to-assets and age in months.[51] As a check against outliers and influential data points,[52] I winsorize all independent variables at their 0.5% and 99.5% values, as in Baker and Wurgler (2006).

To address multicollinearity problems revealed by regression diagnostics, I performed principal component analysis. Principal component analysis derives alternative independent variables that a researcher may put into a regression equation by giving the pre-existing variables different weights and generating a "blended" explanatory variable that the researcher may use in the regression instead. In this analysis, I generated two new variables using this methodology: 1) "retail trading," which combines buy-side retail trading and sell-side retail trading and 2) "size, volume and research," which combines market capitalization, trading volume, research coverage and research activity. In addition, my model excludes 1) the profitability dummy because it is highly correlated with return on equity, 2) the dividend payer dummy because it is highly correlated with dividends to equity and 3) all the sales growth dummies because they are highly correlated with the sales growth variable.[53]

My regression takes the form:

$$\ln(R_i^2/(1-R_i^2)) = \alpha + \beta RETTRADE_i + \gamma_1 I_i + \gamma_2 SIZEVOLRES_i + \gamma_3 INSOWN_i + \gamma_4 5PEROWN_i + \gamma_5 DIVERSE_i + \gamma_6 NEWSDAYS_i + \gamma_7 AGE_i + \gamma_8 STKVOL_i + \gamma_9 RETONEQ_i + \gamma_{10} DIVTOEQ_i + \gamma_{11} TANGASSETS_i + \gamma_{12} RDTOASSETS_i + \gamma_{13} BKTOMKT_i + \gamma_{14} EXTFIN_i + \gamma_{15} SALESG_i + \varepsilon_i \quad (3)$$

where *RETTRADE* is the variable representing the proportion of trading by retail investors, *I* is institutional ownership, *SIZEVOLRES* represents a combined variable including size, trading volume, research coverage and research activity, *INSOWN* is insider ownership, *5PEROWN* is five percent owner ownership, *DIVERSE* is a dummy variable that takes the value one for firms that operate in more than two

[51] In addition, for the variables insider ownership, dividends-to-equity, and R&D-to-assets, I replace all zero values with a small constant (0.001) to allow the calculation of the natural log. However, with or without the addition of this constant, my results are qualitatively the same.

[52] Through the regression diagnostics process, I learned that two of the firms in my sample consistently have values that are outliers. Performing the regression analysis without these two firms causes no qualitative changes in my results, as described in this Section 4.3.

[53] I run one regression analysis using the new variables as described in this paragraph and another that includes each independent variable separately. In unreported results, I find that the results are qualitatively identical.

industries, as indicated by four-digit SIC codes, *NEWSDAYS* is the number of days during the study period on which a firm is prominently featured in a news story, *AGE* is age, in months, *STKVOL* represents the volatility of a firm's stock, *RETONEQ* is return on equity, *DIVTOEQ* is dividends to equity, *TANGASSETS* is the ratio of tangible assets to total assets, *RDTOASSETS* is the ratio of R&D expenses to total assets, *BKTOMKT* is book-to-market equity, *EXTFIN* is level of external financing, and *SALESG* is sales growth.

Table 3 reports the results of this regression, including t-statistics and robust standard errors, for the overall sample and shows that higher proportions of retail trading activity are positively associated with firm-specific return variation (that is, a lower R^2). This result is statistically significant (p-value < 0.01).

Because of the controversy with respect to the correct interpretation of R^2, one must approach the above results with some caution. However, if a low R^2 is a sign of share price accuracy, then the results demonstrate that increased levels of retail trading are associated with more accurate share prices.

4.3.1 Industry Considerations

Some firms move more closely with the market because they are more sensitive to general economic conditions (see Durnev et al. [2003] for a general discussion of this point). Thus, it would not be fair to say that the stock prices of such firms are "less accurate" because market-wide factors largely drive their fundamentals (for example, earnings). To address this potential concern, in unreported results, I perform separate intra-industry regressions of the type described above using a sample of firms in industries that are particularly sensitive to macroeconomic factors (for example, construction companies, finance companies). My overall results still hold qualitatively. Although certain firms are more sensitive to market factors than others, my results reveal that, within groups of such "sensitive" firms, the level of retail trading has a statistically significant negative relationship with R^2.

4.3.2 Size Effects

To assess whether firm size matters with respect to R^2, I perform additional analyses. Firm size seems to be an important consideration in the analysis. Table 1 reveals that, on average, the largest firms in the sample (the "Top Half") have higher R^2's than the smallest firms in the sample (the "Bottom Half"). At first blush, this is counterintuitive. If the prevailing view among R^2 adherents is correct, then this finding implies that larger NYSE firms, as a group, have less accurate stock prices than smaller firms. However, recall the arguments of Lee and Liu (2007) and Dasgupta, Gan and Gao (2010) with respect to R^2. The larger firms relative to the smaller firms are more likely to operate in information-rich environments. Thus, consistent with the claims of Dasgupta, Gan and Gao (2010), these firms, as a group, appear to have less firm-specific information impounded into their prices during the study period, likely because the market largely had anticipated firm-specific news from these firms.

I also find that retail trading in firms of different sizes is associated with R^2 in different ways. I perform separate regressions including first, only the largest firms in the sample (the "Top Half") and then only the smallest firms in the sample (the "Bottom Half").

The results in Table 3 reveal that higher levels of retail trading are associated with lower R^2's in the overall sample and in the Top Half and Bottom Half size groups. However, there is no statistically significant relationship between retail trading activity and R^2 in the Top Half size group. This outcome demonstrates that the relationship between retail trading and R^2 is stronger for relatively smaller[54] firms.

4.3.3 Causation

Though the results described above demonstrate that retail investor trading is correlated with firm-specific return variation, I have not established that retail investor trading *causes* changes in firm-specific return variation. There may be no causal link or the causation may run in the opposite direction. For example, individual investors may be attracted to firms with lower R^2's, and the presence of such investors may have no effect on a stock's R^2. Retail investors that are trying to "beat the market" may be more inclined to invest in stocks that have high firm-specific variation, or stockbrokers' recommendations to their individual investor clients also may tend to consist largely of stocks that have experienced recent movement due to idiosyncratic factors. Finally, firms with greater firm-specific information may garner

[54] I use the term "relatively smaller" because all firms in the study are NYSE firms and among the largest corporations in the United States.

more publicity and, thus, attract more retail investors. Consistent with that hypothesis, Barber and Odean (2008) provide evidence that individual investors are attracted to stocks that "catch the attention" of such investors through extreme price moves, abnormal trading volume and, as noted previously, news coverage.[55] Determining the existence and direction of causation is important for interpreting the results of this study.

There is no way to know with certainty whether retail trading causes changes in R^2. But instrumental variable (IV) estimation is widely used by econometricians to determine the existence of causal relationships and address simultaneity concerns. In this study, through the two-stage least squares method of IV estimation, I, in the first stage, predict retail trading by using a factor (instrument) that is not directly related to R^2. I then, in the second stage, use an IV estimator and the first stage "predicted" results to estimate the effect of retail trading on R^2. "Stock price" is the instrument I use for retail trading. "Stock price" is equal to a firm's average stock price during the study period. Like the independent variables used in the regressions in this study, average stock price in the IV estimation is winsorized at its 0.5% and 99.5% value.

To be a valid instrument, stock price must be correlated with (i) the proportion of retail trading in a stock *(RETTRADE)* and (ii) R^2, but only indirectly through the proportion of retail trading *(RETTRADE)*. The instrument also must be sufficiently strong (that is, have a high correlation with the independent variable of interest, *RETTRADE*). Grullon, Kanatas, and Weston (2004) suggest that individual investors may prefer stocks trading within certain price ranges because of cost concerns. One can imagine an individual investor of relatively modest means that prefers to purchase stock in round lots.[56] To this investor, a stock with a price of $9 may be significantly more attractive than one trading at

[55] It should be noted that, as shown on Table 2, there is not a positive correlation between the level of retail trading and news coverage. Level of a firm's media attention is, in part, a function of size (the correlation between news coverage and market capitalization is 0.58), and size is inversely correlated with the proportion of trading in a firm's stock by retail investors. However, in unreported regression results, I find strong evidence that as the level of news coverage increases, the level of retail trading increases, holding size constant.
[56] The reason for a round lot preference likely is not rooted in concerns about transaction cost differentials. Angel (1997, 62) notes that the odd-lot differential (that is, higher execution costs for odd lot purchases and sales) has been eliminated and that some investors may pay a flat fee per trade, but states, "[n]evertheless, many investors are still

$90 simply because the former is viewed as more affordable (at a total cost per round lot of $900 for the former and $9,000 for the latter). Similarly, if an investor desires diversification (assuming she relies on direct investment rather than investment through intermediaries such as mutual funds for this purpose) and has limited funds available for investment, she may have little choice other than to buy stocks with lower absolute prices to achieve her diversification goals (see Dhar et al. 2004). Grullon, Kanatas, and Weston (2004) employ the reciprocal of stock price as a control variable in a regression assessing the effect of advertising on stock ownership and find a statistically significant relationship (at the 1% level) between stock price and the absolute number of total investors in a firm.[57]

The results of the IV analysis in this study for the overall sample appear in Table 4. Consistent with the findings and hypothesis of Grullon, Kanatas, and Weston, the relationship between the proportion of retail trading and stock price is negative. The first stage results suggest that stock price is a strong instrument for retail trading (F-statistic = 47.81). In addition, the correlation between retail trading and stock price (-0.45) is strong.

In addition to statistical evidence suggesting that stock price is a strong instrument, qualitative reasons to believe the instrument is valid exist. The absolute level of stock price is only indirectly correlated with R^2. Stock price level is unlikely to be directly correlated with R^2 because absolute stock price level should have no effect on share price accuracy or firm-specific return variation. Whether a stock trades at $9 or $90 is an irrelevant consideration with respect to how informative that price is or whether the stock's returns are correlated with the overall market or the firm's industry group. Absolute stock price is an arbitrary figure, devoid of informational content. Though the dependent variable in the main regressions in this study (R^2) is derived from a regression whose dependent variable is stock price

---

reluctant to trade in odd lots." Similarly, Dhar et al. (2004, 19), which examine trading behavior around stock splits, note that though the difference in costs for odd-lot trading and round-lot trading are insignificant during their sample study period (1991-1996), "individual investors tend to like trading in hundreds of shares" and further state that approximately 82% of all common stock trades are round-lot trades.

[57] Note, however, that they fail to find a statistically significant relationship between stock price and the absolute number of institutional investors in the firm.

returns, returns, which reflect relative stock price movements, are independent of absolute stock price levels.

The following simple example illustrates this point. Imagine two identical (with the exception of stock price) firms, each with a current market capitalization of $500 million. Firm A's stock price is $10 per share, and it has 50 million shares outstanding. Firm B, with 25 million shares outstanding, has a stock price of $20. If both Firm A and Firm B experience a negative profitability shock that changes the market's estimate of the firms' value from $500 million to $450 million, the stock prices of both firms will decline by 10% (Firm A's to $9 per share and Firm B's to $18 per share). Though the "pre-shock" and "post-shock" prices are different for Firm A and Firm B, the percentage decline (the return) is the same. Prices and returns are independent, so stock price is likely a valid instrument.

One may argue that this conclusion is not free from doubt because there is some evidence that returns can be related to absolute stock price level. For example, Gaunt, Gray and McIvor (2003, 33), in a study on Australian equity market returns, find that share prices, independent of firm size, affect portfolio returns. Similarly, Bhardwaj and Brooks (1992, 553) find a low price stock January effect (that is, the stocks earn abnormal returns in January).[58] Bhardwaj and Brooks characterize this result as a low price phenomenon because they, like Gaunt, Gray and McIvor, find that the return effects appear generally in stocks with low prices. Bhardwaj and Brooks (1992, 559) suggest that arguments used in the past to explain the anomalous returns of small firms (for example, illiquidity, inaccurate risk assessment, neglect, and transaction costs) can be applied with at least as much force to low price stocks. There is little reason to suspect that the low price phenomenon would affect the results of this study significantly, however. This sample contains firms with relatively high share prices, as the median stock price is $32.89.[59]

---

[58] Note that these abnormal returns disappear when factoring in transaction costs and bid-ask bias to returns in the 1977-1986 period.

[59] Bhardwaj and Brooks (1992, 556) use the following five groups to segment their sample by stock price: $5 or less, $5 - $10, $10 - $15, $15 - $20, and more than $20. Under this construction, stocks whose prices exceed $20 clearly are not low-priced. In addition, the median August 2006 (the final month of this study's analysis period) month-end stock price for the entire CRSP database is $17.09 (excluding firms with "$0" reported stock prices). Approximately 56.7% of these firms have stock prices below $20. Thus, it is reasonable to conclude that the stocks

Therefore, the available evidence suggests that stock price is a valid instrument for retail trading in this context.[60]

The second stage regression demonstrates the effect retail trading has on R^2. This regression shows how R^2 varies with conditions (higher or lower stock prices) that tend to be associated with lower or higher levels of retail trading. As shown in Table 4, the coefficient for retail trading is negative; as the proportion of retail trading across firms in the sample increases, R^2 decreases. The retail trading coefficient in the instrumental variable estimation is statistically significant. The results of IV estimation analysis using the Top Half and Bottom Half size groups, also shown in Table 4, are qualitatively identical to those for the overall sample.[61]

If the assumptions described in this sub-section are correct, the stock price could not cause a firm to have low or high R^2; stock price has no direct effect on R^2. Thus, the results of this analysis provide evidence that the level of a firm's R^2 is caused at least in part by the effect of stock price on the level of retail trading and, in turn, by the effect of retail trading on R^2. Though causality may run in both directions, through the use of this source of exogenous variation (the stock price, which causes one variable (retail trading) to change but not the other (R^2)), I am able to determine the existence of one of the directions of causation and infer the nature and strength of the relationship between retail trading and R^2. This analysis suggests that retail trading causes changes in firm-specific return variation.

4.4 Study Limitations

This study suffers from two limitations. First, it relies on a sample that includes only a subset of publicly traded firms – NYSE-listed companies only. The sample does not include the broader universe

---

used in this study, with an overall median price of \$32.89 and 865 firms (76.6% of the sample) with a price of more than \$20 and only 28 firms (2.5%) with a stock price of under \$5, are not, on the whole, low-priced stocks.

[60] One could argue that low-stock price companies are inherently more volatile, which could result in lower R^2's for such companies. Even if this were true, it would not affect the utility of stock price as an instrument. In this instance, just as described above, stock price still would be correlated with R^2 only indirectly, as it is not the price itself, but rather characteristics of firms with low prices that are correlated with volatility. In addition, just as discussed in note 59 above, few of the firms in this study's sample are low-priced stocks.

[61] Unlike in the OLS regressions, here in the IV estimation, the retail trading coefficient is significant in the Top Half size group. The coefficients for all size groups also are significantly larger (on an absolute basis) in the second stage of the IV estimation than in OLS. This is likely due to the potential reverse causation problem identified in this section. However, just as in the OLS, the relationship between retail trading and R^2 is stronger in the Bottom Half size group than in the Top Half size group.

of firms traded on other exchanges or markets. This is unavoidable given the lack, for non-NYSE companies, of direct (non-proxy), market-wide retail trading data of the sort used in this study. However, using this sample of NYSE firms allows one to test the relationship between R^2 and retail trading among firms that, relative to the broader market, operate in what could be termed a good information environment.

Second, the available data make it impossible to separate the trades of retail investors who are advised by a broker or other financial professional from the trades of individuals that make investment decisions independently.[62] The distinction is meaningful to the extent we hope to draw conclusions about the investment capacity of retail investors to inform securities regulation policy.

5. Discussion and Conclusion

This Article offers evidence that higher levels of trading by retail investors are associated with more firm-specific return variation (lower R^2). The relationship is stronger among the smaller firms in the sample than among the larger firms. This result is reasonable given the differences in retail trading in the larger and smaller firms. As shown in the summary statistics on Table 1, though retail investors on average represent a small proportion of overall stock trading volume, they represent almost twice as much trading volume on a proportional basis for the relatively smaller firms as they do for the relatively larger firms. Thus, for the smaller firms, there is more opportunity for the trades of retail investors to have a meaningful impact on stock price movements.

This Article also provides some evidence that the relationship between retail trading and R^2 is causal. However, because of the controversy surrounding the correct interpretation of R^2 (that is, whether low R^2 is a sign of greater share price accuracy or greater informational inefficiency), the results of this study do not have clear policy implications. If the dominant view among R^2 adherents is correct (that is, lower R^2's imply greater share price accuracy), then the findings of this study suggest that, contrary to the received wisdom, the presence of retail investors, as a group, in equity markets *increases* share price accuracy. Therefore, the instant study not only calls into question the need for policy changes to restrict

[62] Note that some question the value of a broker's advice or her influence on an individual's trading decisions.

the access of retail traders, but also provides support for efforts to protect individual investors and increase their market participation. But if the predominant view is wrong, and a low R^2 is *not* a sign of share price accuracy, but rather a sign of the presence of greater noise in share prices, then the results of this study could lend support to the notion that individual investors are noise traders.

This study is not intended to, nor can it, resolve the controversy surrounding the correct interpretation of R^2. However, it does yield insights that can contribute to the debate. As described in Section 3, Dasgupta, Gan and Gao (2010) argue that the quality of information environment is important in interpreting R^2. They assert that a firm in a high quality information environment is more likely to have a high R^2 during any defined study period because the firm's stock price before the beginning of the study period already reflects the anticipation of firm-specific news. There is less "surprise" and thus less price movement when firm-specific events occur during the study period. Extending the insights of Dasgupta, Gan and Gao (2010),[63] two different types of information are relevant in this context: (1) the "base" of information generally available about a corporation and (2) the "flow" of information during a researcher's study period that affects the incorporation of firm-specific news.

The results[64] of the regressions performed in this study are instructive and reveal that firms that are larger and older, that have higher trading volume and more research coverage, that pay higher dividends relative to book equity, and that have a higher proportion of tangible assets tend to have higher R^2's. All of these factors, as discussed previously in Section 4.3, are likely to contribute to a high quality information environment and make firms operating in this environment less likely to experience non-

[63] Dasgupta, Gan and Gao (2010) use the terms "time-variant" characteristics (that is, those factors reflecting the current state of the firm) and "time-invariant" characteristics (that is, those characteristics that do not change frequently or do not change much over time). One may think of "time-invariant" characteristics as related to the "base" of information or general information environment and "time-variant" characteristics as related to the "flow" of information.

[64] Note that this section describes only those characteristics that have a statistically significant relationship with R^2 in the OLS regressions for the overall sample. Also, note that though sales growth is also a variable that is positively associated with R^2, the meaning of this result is ambiguous. Baker and Wurgler (2006) suggest that sales growth can spur more speculation and hence may be associated with less accurate prices. However, sales growth is also a sign that a firm is not distressed, so arbitrage may be easier. This, in turn, leads to a likelihood of more accurate prices.

market-related stock movements during the study.[65] This study also reveals that firms with higher ownership by 5% holders, higher levels of R&D, and higher book-to-market ratios tend to have lower R^2's. The presence of these characteristics, as discussed previously, is often associated with a poor quality information environment.

All of the traits mentioned in the prior paragraph bear a statistically significant relationship with R^2, either positive or negative, and this is a result that appears to be consistent with the hypothesis of Dasgupta, Gan and Gao (2010). Firms with characteristics that are consistent with a high quality information environment tend to have higher R^2's. Conversely, firms with traits associated with low quality information environments tend to have lower R^2's. These factors all relate to the "base" of information about a firm.

As Durnev et al. (2003) note, information about firm fundamentals is incorporated into stock prices in two ways: through a general revaluation of firm value following a public news release and through investor trading activity following the attainment of private information.[66] Therefore, this Article also provides evidence on the relationship between R^2 and two firm characteristics that relate to the "flow" of information into a firm's stock price: dissemination of firm-specific news and retail trading activity.

The results of this study reveal that higher numbers of news days are associated with lower R^2. Even firms in high quality information environments are not completely transparent; the market cannot anticipate fully all future firm-specific events. Therefore, the presence or absence of firm-specific news is important in explaining stock movements. News coverage reflects events affecting the firm during the study period, such as M&A activity, new customers, or new contract awards. Thus, finding that more

[65] Recall the argument of Dasgupta, Gan and Gao (2010). For firms in high quality information environments, as firm-specific events occur over a defined period of time (such as a limited study period), the stock prices of such firms will move principally with the overall market because investors already will have anticipated firm-specific events, and such expectations will be reflected in the pre-study period price.

[66] Traders likely are more influential on the incorporation of fundamentals into stock prices than is the release of news items. Roll (1988) notes that individual firm-specific stock price movements generally are not correlated with an identifiable public dissemination of news. Thus, Roll argues, trading activity based on investor knowledge and beliefs informed by private information or, alternatively, due to noise, are likely more instrumental in stock price movements than public news releases.

firm-specific news is associated with a firm's stock returns tracking the broader market less closely is unsurprising.[67] News releases are a clear example of an item affecting the "flow" of information into stock prices.[68]

As noted above, retail trading levels also affect the "flow" of information in stock prices. A higher proportion of retail trading generally correlates with a greater number of individuals influencing asset prices.[69] This, of course, directly affects the flow of information into stock prices. This study demonstrates that, just as was the case with greater news coverage, higher levels of retail trading on a proportional basis are associated with lower R^2.

Critics of the prevailing view regarding the correct interpretation of R^2 typically point to implausible correlations between firm characteristics such as small size or less research coverage, on one hand, and low R^2, on the other, if low R^2 is indeed a sign of informational efficiency. The preceding analysis, building on the work of Dasgupta, Gan and Gao (2010), takes a step toward explaining these apparent anomalies. Characteristics consistent with high quality information environments (the "base") are associated with high R^2's and characteristics consistent with increased information flow are associated with low R^2's. The above analysis does not prove that low R^2 is a sign of share price informedness, but the evidence on the relationship between an information flow characteristic such as news coverage and R^2 suggests that one cannot rule that interpretation out as a possibility. Nor can one reject the possibility that trading by retail investors (also associated with low R^2) increases share price accuracy.

Though inconsistent with the conventional wisdom, one could construct a plausible account of how the presence of a greater proportion of individual investors increases share price accuracy. Even if individual investors occasionally or even frequently trade based on noise, they still have information that

[67] Of course, if Dasgupta, Gan and Gao (2010) are correct, then the firms whose stock prices will move the most in reaction to news releases are firms operating in relatively low quality (less transparent) information environments.
[68] Though news coverage is primarily a "flow" of information characteristic as it relates to firm-specific events, it is true that some companies (for example, large firms, firms in popular industries) are more apt, holding all else equal, to attract media attention.
[69] Because retail investors generally buy and sell smaller numbers of shares than institutions, for any given level of trading volume, a higher proportion of retail traders in a stock translates into more separate individuals making a judgment on stock price valuation.

can be valuable in helping the market set prices (see, for example, La Blanc and Rachlinski 2005; Manne 2006; Surowiecki 2004). Thus, average securities prices are more accurate when markets are open, not only to a relatively limited group of investment professionals, but also to all who contribute their bit of knowledge, no matter how small. The findings of Jackson (2003) and Choe, Kho and Stulz (2001), described in Section 2, are consistent with this view.

Despite the foregoing, it is possible that low R^2 is not a sign of share price accuracy and that retail investors are associated with greater noise in market prices. However, even if individual investors, as a group, tend to be noise traders, it does not necessarily follow that market efficiency would be enhanced by excluding retail investors. Market efficiency requires not only accurate prices, but also liquidity (Goshen and Parchomovsky 2006). Of course, eliminating large numbers of investors, of any type, from the market would have significant liquidity implications,[70] and there is evidence that retail investors perform unique market functions (see, for example, Kaniel, Saar and Titman (2008) for a discussion of ways in which individuals provide liquidity for institutional investors). However, putting that very significant issue aside, noise trading itself can perform a vital role in the functioning of financial markets and contribute to market liquidity. If the only trades that occurred were those based on relevant information and all traders had access to (and could act on) the same information, no one, other than for liquidity reasons, would have reason to trade (see Black 1986). Thus, noise trading makes markets more liquid as informed traders attempt to exploit inefficiencies in markets caused by noise trading (Black 1986).[71] Noise trading also indirectly aids in price accuracy, as noise traders make it worthwhile for informed traders to acquire and trade on information that ultimately will make share prices more reflective of fundamental value (La Blanc and Rachlinski 2005).[72]

[70] This is especially true with respect to small firm stocks, for which liquidity often is a significant problem.
[71] On the other hand, limits to arbitrage may make informed traders less willing to trade in the face of large amounts of noise trading (De Long et al. 1990, 703) ("The unpredictability of noise traders' beliefs creates a risk in the price of the asset that deters rational arbitrageurs from aggressively betting against them.").
[72] Unfortunately, the way noise traders make this worthwhile for informed traders is by suffering losses in trades against informed traders (LaBlanc and Rachlinski (2005)).

Assuming there is a causal relationship between individual investor market participation and R^2, no matter what the interpretation of R^2,[73] this study provides further evidence that the trading behavior of individual investors plays an important role in market efficiency. Retail trades, even when small as a percentage of total volume, can have significant market effects.

[73] In addition to the interpretations of R^2 primarily described in this Article, it is possible, of course, that the R^2 metric is devoid of meaning (that is, is neither a sign of share price accuracy or increased noise) and is entirely random. Though theoretically possible, because many researchers have found statistically significant relationships between R^2 and a number of different variables, R^2 likely is a meaningful metric.

References

Angel, James J. 1997. "Picking Your Tick: Toward a New Theory of Stock Splits." *Journal of Applied Corporate Finance* 10:59-68.

Ashbaugh-Skaife, Hollis, Joachim Gassen, and Ryan LaFond. 2006. "Does Stock Price Synchronicity Represent Firm-Specific Information? The International Evidence." MIT Sloan Research Paper No. 4551-05. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=768024.

Baker, Malcolm, and Jeffrey Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *Journal of Finance* 61:1645-1680.

Barber, Brad M., and Terrance Odean. 2008. "All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies* 21:785-818.

Barber, Brad M., Terrance Odean, and Ning Zhu. 2005. "Do Noise Traders Move Markets?" http://www.haas.berkeley.edu/finance/Noise.pdf.

Beny, Laura Nyantung. 2005. "Do Insider Trading Laws Matter? Some Preliminary Comparative Evidence." *American Law and Economics Review* 7:144-83.

Bhardwaj, Rabinder K., and Leroy D. Brooks. 1992. "The January Anomaly: Effects of Low Share Price, Transaction Costs, and Bid-Ask Bias." *Journal of Finance* 47:553-75.

Black, Fischer. 1986. "Noise." *Journal of Finance* 16:529-43.

Blume, Lawrence, and David Easley. 2006. "If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets." *Econometrica* 74: 929-66.

Brusa, Jorge, Pu Liu, and Craig Schulman. 2005. "Weekend Effect, 'Reverse' Weekend Effect, and Investor Trading Activities." *Journal of Business, Finance & Accounting* 32:1495-1517.

Chan, Kalok, and Allaudeen Hameed. 2006. "Stock Price Synchronicity and Analyst Coverage in Emerging Markets." *Journal of Financial Economics* 80:115-47.

Chen, Qi, Itay Goldstein, and Wei Jiang. 2007. "Price Informativeness and Investment Sensitivity to Stock Price." *Review of Financial Studies* 20:619-50.

Choe, Hyuk, Bong-Chan Kho, and René M. Stulz. 2001. "Do Domestic Investors Have More Valuable Information About Individual Stocks than Foreign Investors?" NBER Working Paper no. 8073. http://www.nber.org/papers/w8073.

Choi, Stephen. 2000. "Regulating Investors Not Issuers: A Market-Based Proposal." *California Law Review* 88:279-334.

Dasgupta, Sudipto, Jie Gan, and Ning Gao. 2010, forthcoming. "Transparency, Price Informativeness, Stock Return Synchronicity: Theory and Evidence." *Journal of Financial and Quantitative Analysis.* http://www.bm.ust.hk/~jgan/papers/Transparency%20DRAFT_20090416_final.pdf (draft).

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy* 98:703-38.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldman. 1991. "The Survival of Noise Traders in Financial Markets." *Journal of Business* 64:1–19.

Dhar, Ravi, William N. Goetzmann, Shane Shepherd, and Ning Zhu. 2004. "The Impact of Clientele Changes: Evidence from Stock Splits." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=410104.

Donohue, John J. and Justin Wolfers. 2005. "Uses and Abuses of Empirical Evidence in the Death Penalty Debate." Stanford Law Review 58:791-845.

Durnev, Artyom, Randall Morck, and Bernard Yeung. 2001. "Does Firm-Specific Information in Stock Prices Guide Capital Allocation?" Working Paper no. 8093. National Bureau of Economic Research, Cambridge, MA.

Durnev, Artyom, Randall Morck, and Bernard Yeung. 2004. "Value-Enhancing Capital Budgeting and Firm-specific Stock Return Variation." *Journal of Finance* 59:65-105.

Durnev, Artyom, Randall Morck, Bernard Yeung, and Paul Zarowin. 2003. "Does Greater Firm-Specific Return Variation Mean More or Less Informed Stock Pricing?" *Journal of Accounting Research* 41:797-836.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara. 2005. "Factoring Information Into Returns." EFA 2004 Maastricht Meetings Paper no. 4118. http://ssrn.com/abstract=556079.

Fama, Eugene F. 1965. "The Behavior of Stock Market Prices" *Journal of Business* 38:34-105.

Ferrell, Allen. 2003. "Mandated Disclosure and Stock Returns: Evidence from the Over-the-Counter Market." Harvard Law and Economics Discussion Paper No. 453. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=500123.

Fox, Merritt B., Randall Morck, Bernard Yeung, and Artyom Durnev. 2003. "Law, Share Price Accuracy, and Economic Performance: The New Evidence." *Michigan Law Review* 102:331-86.

Friedman, Milton. 1953. " The Case for Flexible Exchange Rates." 157-203 in *Essays in Positive Economics*. Chicago: The University of Chicago Press.

Gaunt, Clive, Philip Gray, and Julie McIvor. 2000. "The Impact of Share Price on Seasonality and Size Anomalies in Australian Equity Returns." *Accounting and Finance* 40:33-50.

Goldstein, Michael, Andriy Shkilko, Bonnie Van Ness, and Robert Van Ness. 2006. "Competition and Consolidation in the Market for NYSE-Listed Securities." http://www.fma.org/SLC/Papers/Competition_and_Consolidation_in_the_Market_for_the_NYSE-listed_securities.pdf.

Goshen, Zohar , and Gideon Parchomovsky. 2006. "The Essential Role of Securities Regulation." *Duke Law Journal* 55:711-81.

Grullon, Gustavo, George Kanatas, and James P. Weston. 2004. "Advertising, Breadth of Ownership, and Liquidity." *Review of Financial Studies* 17:439-61.

Hameed, Allaudeen, Randall Morck, and Bernard Yeung. 2005. "Information Markets, Analysts, and Comovement in Stock Returns." Unpublished Manuscript.

Hirshleifer, David and Siew Hong Teoh. 2003. "Herd Behaviour and Cascading in Capital Markets: A Review and Synthesis." *European Financial Management* 9:25-66.

Hou, Kewei, Lin Peng and Wei Xiong. 2006. "R^2 and Price Inefficiency." Unpublished manuscript.

Hvidkjaer, Soeren. 2006. "Small Trades and the Cross-Section of Stock Returns." Unpublished manuscript.

Jackson, Andrew. 2003. "The Aggregate Behaviour of Individual Investors." http://ssrn.com/abstract=536942.

Jin, Li, and Stewart C. Myers. 2006. "R^2 Around the World: New Theory and New Tests." *Journal of Financial Economics* 79:257-92.

Kaniel, Ron, Gideon Saar, and Sheridan Titman. 2008. "Individual Investor Trading and Stock Returns." *Journal of Finance* 63:273-310.

Kennedy, Peter. 2003. *A Guide to Econometrics*, 5th ed. Cambridge, MA: The MIT Press.

Kogan, Leonid, Stephen A. Ross, Jiang Wang, and Mark M. Westerfield. 2006. "The Price Impact and Survival of Irrational Traders." *Journal of Finance* 61:195-229.

Kumar, Alok, and Charles M. C. Lee. 2006. "Retail Investor Sentiment and Return Comovements." *Journal of Finance* 61:2451-86.

La Blanc, Gregory, and Jeffrey J. Rachlinski. 2005. "In Praise of Investor Irrationality." 542-88 in *The Law and Economics of Irrational Behavior*, edited by Francesco Parisi and Vernon L. Smith. Stanford, CA: Stanford University Press.

Langevoort, Donald C. 2002. "Taming the Animal Spirits of the Stock Market: A Behavioral Approach to Securities Regulation." *Northwestern University Law Review* 97: 135-88.

Lee, Dong Wook and Mark Liu. 2007. "Does More Information in Stock Price Lead to Greater or Smaller Idiosyncratic Return Volatility?" http://papers.ssrn.com/sol3/papers.cfm?abstract_id=887026.

Lux, Thomas. 1995. "Herd Behaviour, Bubbles and Crashes." *The Economic Journal* 105:881-96.

Manne, Henry. 2006. "Remarks on the Lewis & Clark Law School Business Law Forum: Behavioral Analysis of Corporate Law: Instruction or Distraction?" *Lewis & Clark Law Review* 10:169-76.

Morck, Randall, Bernard Yeung, and Wayne Yu. 2000. "The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements?" *Journal of Financial Economics* 58: 215-60.

New York Stock Exchange. 2004. "NYSE Order Tracking System, Input File Layout, v. 4.3." http://www.nyse.com/pdfs/order_tracking_system_v4.3b.pdf.

Patterson, Scott and Aaron Lucchetti. 2008. Boom in 'Dark Pool' Trading Networks Is Causing Headaches on Wall Street. *Wall Street Journal*, May 8.

Piotroski, Joseph D., and Darren T. Roulstone. 2004. "The Influence of Analysts, Institutional Investors, and Insiders on the Incorporation of Market, Industry and Firm-Specific Information into Stock Prices." *The Accounting Review* 79: 1119-51.

Pizarro, Veronica, Sakthi Mahenthiran, David Cademartori and Roberto Curci. 2007. "The Influence of Insiders and Institutional Owners on the Value, Transparency, and Earnings Quality of Chilean Listed Firms?" http://papers.ssrn.com/sol3/papers.cfm?abstract_id=982697.

ReTracEOD Data Discussion Board. http://www.nysedata.com/nysedata/Support/DiscussionBoard/tabid/108/view/topic/postid/28/forumid/6/tpage/1/Default.aspx (last viewed October 18, 2007).

Roll, Richard. 1988. "R^2." *Journal of Finance* 43: 541-66.

Sandroni, Alvaro. 2000. "Do Markets Favor Agents Able to Make Accurate Predictions?" *Econometrica* 68:1303–41.

Surowiecki, James. 2004. *The Wisdom of Crowds*: *Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Doubleday.

Teoh, Siew Hong, Yong Yang, and Yinglei Zhang. 2006. "R-square: Noise or Firm-Specific Information?." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=926948.

Thakor, Anjan V., with Jeffrey S. Nielsen, and David A. Gulley. 2005. "The Economic Reality of Securities Class Action Litigation." U.S. Chamber Institute for Legal Reform.

Tobin, James. 1982. "On the Efficiency of the Financial System." *Lloyd's Banking Review* 153:1-15.

Varadarajan, P. "Rajan", and Vasudevan Ramanujam. 1987. "Diversification and Performance: A Reexamination Using a New Two-Dimensional Conceptualization of Diversity in Firms." *The Academy of Management Journal* 30: 380-93.

Veldkamp, Laura L. 2006. "Information Markets and the Comovement of Asset Prices." *The Review of Economic Studies* 73:823-45.

Winter, Ralph K. 1988. "On 'Protecting the Ordinary Investor'." *Washington Law Review* 63:881-902.

Wu, Hsiu-Kwang. 1972. "Odd-Lot Trading in the Stock Market and Its Market Impact." *The Journal of Financial and Quantitative Analysis* 7:1321-41.

Wurgler, Jeffrey. 2000. "Financial Markets and the Allocation of Capital." *Journal of Financial Economics* 58:187-214.

Table 1

Summary Statistics

Dependent Variable		Overall Sample	Top Half	Bottom Half
R^2	Mean	0.287	0.326	0.247
	Std. Dev.	0.177	0.190	0.153
	Minimum	0.0005	0.036	0.0005
	Maximum	0.856	0.856	0.827
Independent Variables of Interest				
Ratio of Retail Buy-side Shares to Total Daily Volume	Mean	0.011	0.008	0.015
	Std. Dev.	0.011	0.006	0.014
	Minimum	0.001	0.001	0.002
	Maximum	0.099	0.055	0.099
Ratio of Retail Sell-side Shares to Total Daily Volume	Mean	0.017	0.012	0.021
	Std. Dev.	0.015	0.009	0.018
	Minimum	0.002	0.003	0.002
	Maximum	0.117	0.110	0.117
Control Variables				
Institutional Ownership %	Mean	0.722	0.714	0.729
	Std. Dev.	0.216	0.199	0.231
	Minimum	0.020	0.020	0.026
	Maximum	1.723	1.723	1.267
Market Capitalization (mm)	Mean	8,796	16,525	1,054
	Std. Dev.	24,056	32,207	590
	Minimum	79	2,308	79
	Maximum	375,773	375,773	2,299
Total Daily Trading Volume	Mean	1,386,455	2,332,806	438,425
	Std. Dev.	2,786,927	3,664,238	558,481
	Minimum	9,023	14,800	9,023
	Maximum	46,642,493	46,642,493	5,923,069

Research Coverage	Mean	13	18	8
	Std. Dev.	8	8	5
	Minimum	1	2	1
	Maximum	47	47	35
Research Activity	Mean	449	658	240
	Std. Dev.	404	443	210
	Minimum	1	32	1
	Maximum	2,498	2,498	1,674
Insider Ownership	Mean	0.069	0.052	0.086
	Std. Dev.	0.144	0.140	0.145
	Minimum	0	0	0
	Maximum	.9999	.9999	.9999
5% Owner Ownership	Mean	0.353	0.290	0.419
	Std. Dev.	0.223	0.203	0.224
	Minimum	0	0	0
	Maximum	.9999	.9999	.9999
Industry Groups (Four-Digit SIC Codes)	Mean	3	4	3
	Std. Dev.	1.681	1.727	1.578
	Minimum	1	1	1
	Maximum	7	7	7
News Days	Mean	34	47	22
	Std. Dev.	44	58	13
	Minimum	0	3	0
	Maximum	369	369	112
Age (in months)	Mean	311	369	252
	Std. Dev.	247	271	204
	Minimum	15	15	15
	Maximum	951	951	951
Volatility (Standard Deviation of Monthly Returns)	Mean	0.083	0.069	0.098
	Std. Dev.	0.038	0.027	0.041
	Minimum	0.021	0.021	0.022
	Maximum	0.376	0.182	0.376

Return on Equity	Mean	0.145	0.225	0.065
	Std. Dev.	1.043	1.421	0.378
	Minimum	-6.143	-1.661	-6.143
	Maximum	33.255	33.255	1.836
Dividends-to-Equity	Mean	0.055	0.066	0.044
	Std. Dev.	0.649	0.752	0.526
	Minimum	-.870	-.136	-.870
	Maximum	17.823	17.823	12.298
Tangible Assets to Total Assets	Mean	0.565	0.559	0.570
	Std. Dev.	0.396	0.399	0.394
	Minimum	0.000	0.000	0.000
	Maximum	2.709	1.868	2.709
R&D to Assets	Mean	0.027	0.029	0.025
	Std. Dev.	0.037	0.036	0.038
	Minimum	0.000	0.000	0.000
	Maximum	0.367	0.173	0.367
Book-to-Market Equity	Mean	0.491	0.414	0.568
	Std. Dev.	0.446	0.278	0.556
	Minimum	-6.769	-0.454	-6.769
	Maximum	4.092	2.305	4.092
External Financing	Mean	0.040	0.044	0.035
	Std. Dev.	0.385	0.148	0.524
	Minimum	-11.687	-1.623	-11.687
	Maximum	0.768	0.751	0.768
Sales Growth	Mean	0.153	0.152	0.154
	Std. Dev.	0.207	0.194	0.220
	Minimum	-0.903	-0.351	-0.903
	Maximum	1.846	1.846	1.683
Instrumental Variable				
Stock Price	Mean	37.80	48.80	26.77
	Std. Dev.	37.52	47.10	18.81
	Minimum	1.59	2.75	1.59
	Maximum	790.33	790.33	287.43

Table 2
Correlation Matrix for Overall Sample

	Retail Trading	Inst. Ownrshp.	Size, Vol. & Res.	Insider Ownrshp.	5% Owner Ownrshp.	Diverse	News Days	Age	Volatility	Return on Equity	Dividends -to-Equity	Tang. Assets to Total Assets	R&D to Assets	Book-to-Mkt. Equity	External Finance	Sales Growth
Retail Trading	1															
Inst. Ownrshp.	-0.465** (0.000)	1														
Size, Vol. & Res.	-0.442** (0.000)	0.094** (0.002)	1													
Insider Ownrshp.	0.019 (0.527)	0.130** (0.000)	-0.248** (0.000)	1												
5% Owner Ownrshp.	0.073* (0.015)	0.104** (0.001)	-0.309** (0.000)	0.230** (0.000)	1											
Diverse	-0.040 (0.178)	-0.058+ (0.053)	0.078** (0.008)	-0.064* (0.032)	-0.106** (0.000)	1										
News Days	-0.148** (0.000)	-0.015 (0.611)	0.595** (0.000)	-0.204** (0.000)	-0.208** (0.000)	0.145** (0.000)	1									
Age	-0.115** (0.000)	-0.005 (0.876)	0.129** (0.000)	-0.082** (0.006)	-0.232** (0.000)	0.284** (0.000)	0.131** (0.000)	1								
Volatility	0.360** (0.000)	0.029 (0.334)	-0.247** (0.000)	0.092** (0.002)	0.259** (0.000)	-0.173** (0.000)	-0.107** (0.000)	-0.228** (0.000)	1							
Return on Equity	-0.132** (0.000)	0.042 (0.159)	0.134** (0.000)	0.019 (0.524)	-0.064* (0.031)	-0.038 (0.198)	-0.022 (0.468)	0.024 (0.413)	-0.226** (0.000)	1						
Dividends -to-Equity	-0.099** (0.001)	-0.176** (0.000)	0.149** (0.000)	-0.123** (0.000)	-0.248** (0.000)	(0.175)** (0.000)	0.132** (0.000)	0.368** (0.000)	-0.383** (0.000)	0.148** (0.000)	1					
Tang. Assets to Total Assets	0.094** (0.002)	-0.090** (0.003)	0.040 (0.175)	-0.137** (0.000)	-0.075* (0.011)	0.011 (0.716)	-0.019 (0.531)	0.146** (0.000)	0.104** (0.000)	-0.057+ (0.057)	0.097** (0.001)	1				
R&D to Assets	-0.047 (0.111)	(0.037) (0.212)	0.049+ (0.098)	-0.087** (0.003)	-0.046 (0.124)	0.043 (0.153)	0.097** (0.001)	0.118** (0.000)	0.030 (0.322)	-0.068* (0.023)	(0.006) (0.833)	-0.125** (0.000)	1			
Book-to-Mkt. Equity	0.141** (0.000)	-0.049 (0.101)	-0.275** (0.000)	-0.023 (0.438)	0.094** (0.002)	-0.004 (0.888)	-0.095** (0.002)	-0.019 (0.521)	-0.001 (0.987)	-0.232** (0.000)	0.026 (0.390)	(0.003) (0.918)	-0.151** (0.000)	1		
External Finance	0.095** (0.001)	0.033 (0.275)	0.023 (0.449)	0.043 (0.152)	-0.017 (0.580)	-0.023 (0.445)	-0.015 (0.626)	-0.103** (0.001)	0.097** (0.001)	-0.020 (0.497)	-0.062* (0.037)	-0.071 (0.017)*	-0.051+ (0.086)	-0.081** (0.006)	1	
Sales Growth	0.116** (0.000)	0.016 (0.581)	0.080** (0.007)	0.041 (0.167)	-0.038 (0.203)	-0.067* (0.025)	-0.050+ (0.092)	-0.127** (0.000)	0.161** (0.000)	0.152** (0.000)	-0.101** (0.001)	0.030 (0.320)	-0.016 (0.594)	-0.148** (0.000)	0.482** (0.000)	1

NOTE.—The table above reports pairwise correlation coefficients for the independent variables used in the analysis of the relationship between R^2 and retail trading for the overall sample. The numbers in parentheses signify the probability levels at which one may reject the null hypothesis of zero correlation in two-tailed tests. +, *, and ** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 3
Estimates of the Relationship between R^2 and Retail Trading Activity

	Retail Trading		
	Overall Sample	Top Half	Bottom Half
R^2	(1)	(2)	(3)
Constant	-1.681**	-2.399**	-.916*
	(-5.65)	(-5.90)	(-2.01)
	[.297]	[.407]	[.457]
Retail Trading	-.112**	-.028	-.173**
	(-3.07)	(-0.56)	(-3.39)
	[.036]	[.050]	[.051]
Institutional Ownership	.202	-.039	.175
	(1.07)	(-0.15)	(0.67)
	[.189]	[.258]	[.262]
Size, Volume & Research	.171**	.219**	.197**
	(6.99)	(5.77)	(3.93)
	[.024]	[.038]	[.050]
Insider Ownership	.014	-.002	.033
	(0.95)	(-0.11)	(1.50)
	[.014]	[.018]	[.022]
5% Owner Ownership	-.423**	-.476*	-.416*
	(-3.19)	(-2.46)	(-2.32)
	[.133]	[.193]	[.179]
Diverse	.031	.149+	-.074
	(0.53)	(1.82)	(-0.94)
	[.058]	[.082]	[.079]
News Days	-.207**	-.213**	-.227**
	(-4.14)	(-3.28)	(-2.76)
	[.050]	[.065]	[.082]
Age	.144**	.159**	.174**
	(4.32)	(3.37)	(3.41)
	[.033]	[.047]	[.051]
Volatility	-1.268	3.855*	-3.945*

	(-1.03)	(2.50)	(-2.43)
	[1.223]	[1.540]	[1.621]
Return on Equity	-.189	-.162	-.146
	(-1.33)	(-0.84)	(-0.71)
	[.142]	[.193]	[.207]
Dividends-to-Equity	.039**	.040*	.036*
	(3.16)	(2.42)	(2.09)
	[.012]	[.016]	[.017]
Tangible Assets to	.638**	.615**	.498**
Total Assets	(7.38)	(5.60)	(3.96)
	[.086]	[.110]	[.126]
R&D to Assets	-.062**	-.080**	-.038*
	(-6.09)	(-5.92)	(-2.59)
	[.010]	[.014]	[.015]
Book-to-Market Equity	-.210*	.124	-.381**
	(-2.03)	(0.89)	(-2.96)
	[.103]	[.140]	[.129]
External Finance	.207	.486	-.046
	(0.90)	(1.60)	(-0.15)
	[.229]	[.304]	[.310]
Sales Growth	.594**	.530*	.505$^+$
	(3.18)	(2.25)	(1.83)
	[.187]	[.235]	[.276]
F-Statistic	25.33	18.54	10.43
R^2	0.32	0.33	0.32
No. of observations	1129	565	564

NOTE.—The table above reports ordinary least squares regression results for the overall sample and for segments of the sample determined by size (based on average market capitalization during the study period). The segments are "Top Half" (largest - top 50%) and "Bottom Half" (smallest - bottom 50%). $^+$, *, and ** indicate significance at the 10%, 5% and 1% levels, respectively. Numbers in parentheses are heteroskedasticity-robust t-statistics. Numbers in brackets are robust standard errors.

Table 4

Instrumental Variable Estimates of the Relationship between R^2 and Retail Trading Activity

	Overall Sample		Top Half		Bottom Half	
	First Stage Retail Trading	Second Stage R^2	First Stage Retail Trading	Second Stage R^2	First Stage Retail Trading	Second Stage R^2
Constant	1.249 (3.57) [.350]	-1.788** (-4.26) [.420]	.817$^+$ (1.73) [.474]	-2.545** (-3.32) [.766]	1.628** (3.41) [.478]	-.693 (-1.24) [.558]
Retail Trading		-1.163** (-5.65) [.206]		-1.813** (-2.91) [.622]		-1.076** (-4.46) [.241]
Institutional Ownership	-2.226** (-16.02) [.139]	-2.307** (-4.11) [.561]	-2.178** (-10.23) .213	-4.118** (-2.74) [1.50]	-2.037** (-11.22) [.182]	-1.854** (-2.87) [.647]
Size, Volume & Research	-.257** (-12.47) [.021]	-.136* (-2.05) [.066]	-.139** (-4.19) [.033]	-.034 (-0.31) [.111]	-.425** (-11.34) [.037]	-.213$^+$ (-1.81) [.118]
Insider Ownership	-.012 (-0.98) [.013]	.001 (0.04) [.020]	.003 (0.17) [.017]	.008 (0.22) [.036]	-.025 (-1.42) [.017]	.003 (0.12) [.027]
5% Owner Ownership	-.273* (-2.11) [.130]	-.749** (-3.70) [.202]	-.363$^+$ (-1.85) [.196]	-1.110* (-2.40) [.461]	-.206 (-1.23) [.167]	-.674** (-2.73) [.246]
Diverse	.022 (0.41) [.055]	.041 (0.51) [.081]	.106 (1.41) [.075]	.325$^+$ (1.92) [.169]	-.021 (-0.27) [.077]	-.092 (-0.91) [.101]
News Days	.182** (4.26) [.043]	.019 (0.24) [.079]	.227** (4.15) [.055]	.216 (1.17) [.184]	-.021 (-0.30) [.070]	-.218* (-2.21) [.099]

Age	-.009	.132**	-.043	.092	-.032	.133*
	(-0.26)	(2.86)	(-0.98)	(1.03)	(-0.66)	(2.14)
	[.033]	[.046]	[.044]	[.090]	[.048]	[.062]
Volatility	6.77**	7.729**	6.712**	16.778**	6.638**	3.859
	(7.49)	(3.06)	(4.88)	(3.00)	(5.85)	(1.26)
	[.905]	[2.526]	[1.376]	[5.601]	[1.134]	[3.074]
Return on Equity	.134	-.169	-.211	-.626+	.205	-.086
	(1.09)	(-0.95)	(-1.43)	(-1.93)	(1.29)	(-0.38)
	[.124]	[.178]	[.148]	[.325]	[.159]	[.226]
Dividends-to-Equity	.004	.023	.023	.061+	-.016	.003
	(0.35)	(1.40)	(1.47)	(1.83)	(-1.12)	(0.14)
	[.011]	[.016]	[.016]	[.033]	[.014]	[.022]
Tangible Assets to Total Assets	.107	.781**	.108	.855**	.069	.573**
	(1.55)	(6.87)	(1.15)	(3.97)	(0.70)	(3.88)
	[.069]	[.114]	[.094]	[.215]	[.099]	[.148]
R&D to Assets	-.015*	-.071**	-.036**	-.138**	-.006	-.036+
	(-1.52)	(-4.94)	(-2.80)	(-4.32)	(-0.41)	(-1.92)
	[.010]	[.014]	[.013]	[.032]	[.015]	[.019]
Book-to-Market Equity	-.030	-.088	-.406**	-.472	.053	-.181
	(-0.36)	(-0.75)	(-3.06)	(-1.48)	(0.55)	(-1.27)
	[.081]	[.117]	[.133]	[.319]	[.095]	[.142]
External Finance	.398*	.640*	.739*	1.852*	.279	.196
	(2.07)	(1.97)	(2.57)	(2.40)	(1.22)	(0.51)
	[.193]	[.325]	[.288]	[.772]	[.228]	[.381]
Sales Growth	.799**	1.200**	.953**	1.967**	.774**	.988**
	(4.77)	(4.27)	(3.70)	(2.80)	(3.72)	(2.81)
	[.168]	[.281]	[.258]	[.702]	[.208]	[.351]
Stock Price	-.342**		-.226**		-.378**	
	(-6.91)		(-3.15)		(-5.53)	
	[.049]		[.072]		[.068]	
F-Statistic	47.81		9.91		30.53	

No. of observations:	1129	565	564

NOTE.—The table above reports the results of a regression of R^2 on retail trading for the overall sample and for segments of the sample determined by size (based on the average market capitalization during the study period), estimated using 2SLS (two-stage least squares regression). The segments are "Top Half" (largest – top 50%) and "Bottom Half" (smallest – bottom 50%). Stock price is used as an instrument for retail trading. Numbers in parentheses are heteroskedasticity-robust t-statistics or z-scores. Numbers in brackets are robust standard errors. $^+$, *, and ** indicate significance at the 10%, 5% and 1% levels, respectively.