2014

Market Efficiency and the Problem of Retail Flight

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

Recommended Citation

Davis, Alicia J. "Market Efficiency and the Problem of Retail Flight." Stan. J. L. Bus. & Fin. 20, no. 1 (2014): 36-90.

Market Efficiency and the Problem of Retail Flight

Alicia J. Davis*

In 1950, 91% of common stock in the U.S. was owned directly by individual investors. Today, that percentage stands at only 23%. The mass exodus of retail investors and their investment dollars has negative implications not only for capital formation and investor protection, but also for market efficiency. Individual investors are often assumed to be noise traders who distort stock prices and harm market functioning. Therefore, some argue that their withdrawal from the market should be of little concern; indeed, it should be celebrated. Recent empirical evidence calls this assertion of retail noise trading into doubt, and this paper, which describes a study that employs New York Stock Exchange retail trading data, contributes to the debate. This study (1) reveals that as the proportion of trading by individual investors increases, stock price informativeness, as measured by firm-specific return variation (R^2) and the probability of informed trade (PIN), increases and (2) provides evidence that suggests these relationships are causal ones. This study, therefore, provides evidence that, contrary to the received wisdom, retail trading may increase share price accuracy and market efficiency. Thus, there may be substantial reasons to lament retail investor flight.

---

* Professor, University of Michigan Law School. For helpful conversations and comments on earlier drafts, I thank Steve Choi, John DiNardo, Mitu Gulati, Todd Henderson, Don Herzog, Jim Hines, Vic Khanna, Saul Levmore, Nina Mendelson, Martha Nussbaum, JJ Prescott, Adam Pritchard, Veronica Santarosa, Sonja Starr and participants at the Conference on Empirical Legal Studies, the University of Chicago Law School's Best Ideas lecture series, the Annual Meeting of the American Law and Economics Association, and the University of Michigan Law School Law and Economics Workshop. For research assistance, I thank Zachary Klausner and Jesse Hogin, and for administrative assistance, I thank Karen Rushlow. For statistical consulting services on an earlier, related project, 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 and Michael Ellenbogen, Anthony Rubin, and Caryn Smith for research assistance on an earlier, related project. Any errors are my own. The Cook Fund of the University of Michigan Law School and the Walter V. Schaefer Fund at the University of Chicago Law School provided financial support for this project.

I. Introduction

Retail investors are abandoning the U.S. equity markets. In 1950, individuals owned 91% of stocks in the U.S.[1] In 2005, such investors owned 32% of U.S. stocks.[2] Today, that number stands at 23%.[3] The fact that there has been a mass exodus of individuals as direct participants in the stock market is not in dispute. The question we as a society face today is whether we should care. This Article provides evidence that suggests we should.

The Securities and Exchange Commission (SEC) has a three-part mission: the protection of investors; the facilitation of capital formation; and the maintenance of fair, orderly and efficient markets.[4] From a capital formation perspective, direct retail divestment is arguably of little concern if retail investors are shifting their investment dollars to intermediaries investing on their behalves. It is true that while direct investment in the stock market by individual investors is down, indirect ownership (e.g., retail investors holding shares indirectly through a mutual fund intermediary or pension plans investing employer contributions on behalf of individuals) has increased exponentially over the last several decades. For example, in 1950, when direct retail stock ownership was 91%, mutual fund stock ownership was just 3%, and pension plan ownership was only 1%.[5] Today, while retail ownership is 23%, institutional investors[6] own approximately 60% of the stocks in the public equity markets.[7] However, the migration of direct investment dollars to intermediaries is incomplete.

The following chart illustrates equity market inflows and outflows from 2009-2013E:

---

[1] John C. Bogle, Editorial, *Individual Stockholder, R.I.P.*, WALL ST. J., Oct. 3, 2005, at A16.

[2] *Id.*

[3] GOLDMAN SACHS PORTFOLIO STRATEGY RESEARCH, GOLDMAN SACHS, AN EQUITY INVESTOR'S GUIDE TO THE FLOW OF FUNDS ACCOUNTS 5 (Mar. 11, 2013).

[4] SEC. & EXCH. COMM'N, THE INVESTOR'S ADVOCATE: HOW THE SEC PROTECTS INVESTORS, MAINTAINS MARKET INTEGRITY, AND FACILITATES CAPITAL FORMATION (2014), *available at* http://www.sec.gov/about/whatwedo.shtml.

[5] Bogle, *supra* note 1.

[6] Defined here as mutual funds, pension plans, insurance companies, exchange-traded funds, and hedge funds

[7] GOLDMAN SACHS PORTFOLIO STRATEGY RESEARCH, *supra* note 3, at 5. The remaining equity is held by non-US investors (9%), individuals with large stakes, trusts and endowments (4%), and other (e.g., banks, broker-dealers, federal, state and local government) (5%). Total does not add to 100% due to rounding. *Id.*

U.S. Net Equity Flows
(Dollars in billions)

Sector	**2009**	**2010**	**2011**	**2012**	**2013E**
Households	$59	($95)	($62)	($204)	($475)
Mutual Funds	$86	$44	$5	($37)	$125
Life Insurance	$33	$46	$38	$40	$50
Pension Plans	($155)	($135)	($130)	($84)	($100)
Net Household-Related Inflow/(Outflow)	**$17**	**($140)**	**($149)**	**($285)**	**($400)**

Data Source: Federal Reserve Board Flow of Funds Data and Goldman Sachs Global ECS Research[8]

As shown above, the U.S. household sector removed an estimated *$475 billion* from the U.S. equity capital markets in 2013 alone.[9] Though some of the household sector net outflow was offset by estimated net inflows from mutual funds ($125 billion) and life insurance holdings ($50 billion), the vast majority of household equity outflows did not make their way back into equity markets through other means. Indeed, pension plans, whose funding includes direct contributions from individuals (i.e., defined contribution plans such as 401(k)'s) and contributions on their behalves by private or government employers (i.e., defined benefit plans), withdrew an estimated net $100 billion from equity markets in 2013. All total, 2013 saw an estimated net outflow of $400 billion from the household sector. This is 40% more than the $285 billion net outflow in 2012, 168% more than the $285 billion net outflow in 2011, and 186% more than the $140 billion net outflow in 2010. Since 2009, net outflows from the retail sector, *after* accounting for the equity inflows/outflows of intermediaries, total almost *$1 trillion*. So, it is clear that direct retail investment dollars are not being replaced by indirect investments.

Moreover, according to an April 2013 Gallup Poll, only 52% of Americans own stocks either directly or through an intermediary such as a mutual fund or self-

[8] *Id.* at 1.

[9] *Id.* The household sector data used in this Article come from the Federal Reserve Board's Flow of Funds Report. The "household sector" is a residual category and includes data that represents not only retail investors, but also nonprofits, endowments, hedge funds, and private equity funds. However, Goldman Research estimates that approximately 60.5% of the stock ownership attributed to "households" can be attributed to retail investors. Thus, it is reasonable to conclude that retail investors compose the overwhelming majority of these outflows, especially in light of the economic condition of individuals following the Financial Crisis. (See discussion *infra*.)

directed retirement account.[10] This is the lowest level of American stock ownership since Gallup began tracking stock ownership in 1998 and is (1) 10 percentage points lower than the 62% of Americans who were invested in the stock market pre-financial crisis in 2008 and (2) 15 percentage points lower than the high of 67% in 2002.[11] This survey asks if individuals have stocks invested either directly *or indirectly* in the stock market. Therefore, if there were just a shifting of dollars from direct investment to investment through intermediaries, we would not see the decline in ownership percentages over time. This suggests that retail investors are abandoning the stock market altogether.

The retail equity market disinvestment trend is not expected to reverse, but instead accelerate for a number of reasons, including 1) an aging population less likely to invest in the stock market,[12] 2) rising allocations to alternative investments (e.g., private equity, hedge funds and real estate) by wealthy individuals seeking higher returns than those available in public equity markets,[13] 3) declining confidence in U.S. markets by American citizens,[14] and 4) less disposable income available for investment due to high unemployment and a slow economic recovery for the middle class.[15] With respect to the last point, the 2013 Gallup Poll cited above reveals that while overall stock market ownership fell from 62% of Americans to 52% of Ameri-

---

[10] *Only Half of All Americans Invested in Stocks*, CNN MONEY (May 9, 2013), http://money.cnn.com/2013/05/09/investing/american-stock-ownership/index.html.

[11] Lydia Saad, *U.S. Stock Ownership Stays at Record Low*, GALLUP.COM (MAY 8, 2013), http://www.gallup.com/poll/162353/stock-ownership-stays-record-low.aspx. (Click on "View methodology, full question results, and trend data." at the bottom of the page to view historical chart.)

[12] CHARLES ROXBURGH ET AL., MCKINSEY GLOBAL INSTITUTE, THE EMERGING EQUITY GAP: GROWTH AND STABILITY IN THE NEW INVESTOR LANDSCAPE 5 (2011) ("As investors enter retirement, they typically stop accumulating assets and begin to rely on investment income; they shift assets from equities to bank deposits and fixed-income instruments. This pattern has led to predictions of an equity sell-off as the enormous baby boom generation in the United States . . . enters retirement (the oldest members of this cohort reached 65 in 2011) . . . ."). McKinsey feels as though the fear of an equity divestment may be overblown, but does acknowledge that "if investors retiring in the next ten years maintain the equity allocations of today's retirees, equities will fall from 42 percent of US household portfolios to 40 percent in 2020—and to even 38 percent by 2030." *Id; see also* Hibah Yousuf, *Investors Yank $150 Billion from Stocks for 3rd Year*, CNN MONEY (Dec. 27, 2012), http://buzz.money.cnn.com/2012/12/27/investors-stocks-bonds/?iid=EL ("Experts have largely pinned the reason for the stock dump on the Baby Boomer generation. They represent the largest group among retail investors, and after having their portfolios rocked by the dot-com crash and the financial crisis, they've shifted out of stocks and into bonds much earlier than usual as they head into retirement. And any money that they do have in the market is consistently going into bond funds . . . .")

[13] ROXBURGH, *supra* note 12, at 5.

[14] *Id; see also* Yousuf, *supra* note 12 ("A lack of confidence among investors across age groups has also been a factor, in the wake of high-frequency trading, and incidents like the May 2010 flash crash, Nasdaq's bungled Facebook IPO (FB) and the Knight trading glitch (KCG).").

[15] *Only Half of All Americans Invested in Stocks*, *supra* note 10.

cans in the five years since the start of the financial crisis, the biggest decline in stock ownership occurred in the middle class.[16] Among those earning $75,000 or more per year, stock ownership fell from 88% to 81%, a change of seven percentage points.[17] Among those earning between $35,000 and $74,999, however, stock ownership declined by 16 percentage points, from 66% to 50%.[18] As the above chart shows, in a mere three years, household sector outflows ballooned from $95 billion/year to $475 billion/year, a 400% increase. If this trend continues, the implications for capital formation are significant. If retail investors continue to withdraw such large amounts of capital from the market, it can have deleterious effects on issuers. Companies need equity capital for growth and to provide a cushion in times of stress that debt financing cannot provide.[19] Retail disinvestment is particularly problematic for small capitalization companies that have a difficult time attracting the investment dollars of large institutions.[20]

Second, from the perspective of investor protection, there is reason to be concerned about retail flight. Jasmin Sethi argues that securities regulators should move beyond merely protecting individual investors from harm; they should extend their mandate to include promoting market participation and individual saving.[21] As noted above, the equity markets experienced a mass retail investor exodus from 2009-2013. The dramatic increase in outflows in 2013 happened in a year when stock market returns were at record levels. The S&P 500 was up 31% in 2013, almost doubling the 16% return achieved in 2012.[22] During the five-year period when the household-related sector was withdrawing almost $1 trillion (net)[23] from equity

---

[16] Saad, *supra* note 11.

[17] *Id.*

[18] *Id.*

[19] ROXBURGH, *supra* note 12, at 8.

[20] Small capitalization companies rely heavily on retail sector investment because of limitations (e.g., size, ownership, execution risk) on the ability of institutional investors to invest in small cap companies. Alicia Davis Evans, *A Requiem for the Retail Investor?*, 95 VA. L. REV. 1105, 1117 (2009).

[21] *See* Jasmin Sethi, *Another Role for Securities Regulation: Expanding Investor Opportunity*, 16 FORDHAM J. CORP. & FIN. L. 783 (2011) (arguing that one goal of securities regulation should be expanding opportunities for wealth accumulation across various sectors of the U.S. population).

[22] Robert Lenzner, *The Argument for a Sixth Year of Bull Market*, FORBES (Dec. 30, 2013), http://www.forbes.com/sites/robertlenzner/2013/12/30/i-dont-know-anyone-who-predicted-a-five-year-bull-market/. Not that the pace slowed in the first half of 2014, when the S&P increased 6.1%. *See* Angela Moon, *S&P 500, NASDAQ Score Sixth Straight Quarter of Gains*, REUTERS (Jun. 30, 2014), *available at* http://www.reuters.com/article/2014/06/30/us-markets-stocks-idUSKBN0F119U20140630.

[23] Includes net household-related outflows of $400 billion in 2013, $285 billion in 2012, $149 billion in 2011, $140 billion in 2010, plus net inflows of $17 billion in 2009 total $957 billion.

markets, the S&P 500 almost tripled.[24]

Retail investors were sitting on the sidelines while those invested in the equity capital markets have reaped large rewards. The typical retail investor does not have access to high-return investment opportunities such as private equity, hedge funds or real estate, and, thus, are generally left with public equity or debt as investment options. There is some evidence that retail investors fleeing equity capital markets turned to the debt markets.[25] However, debt, though it can be a part of well-balanced investment portfolio, is a poor substitute for equity. The return on debt products has been at historical lows, with interest rates on savings accounts averaging 0.06% annually in 2013, with many of the largest banks in the U.S. paying only 0.01%.[26] The average bond fund had a return of -2.0% in 2013.[27] Investors in the bond market lost money, while equity investors earned a 31% return. Over the long-term, despite occasional market downturns, equity securities provide a higher return than debt securities. Equity investment is an important part of an investment plan for those hoping to meet their savings goals for retirement[28] or life events (e.g., college, first home), so there are significant reasons to encourage retail investors to return to the equity capital markets.

In this Article, I argue that there is another seemingly counterintuitive reason to worry about retail flight - its effect on market efficiency. Market efficiency requires liquidity and accurate prices.[29] The exit of large numbers of investors, of any type, from the market has significant liquidity implications, so retail divestment harms market efficiency in this way. This is especially worrisome with respect to small firm stocks, for which liquidity often is a significant problem. Moreover, there is evidence that retail investors perform unique market functions, as they provide liquidity to institutional investors who require immediacy in trade execution.[30]

I further submit that we would have reason to be concerned about direct retail divestment even if 100% of direct retail investment dollars migrated to intermediaries. The loss of direct retail participation - not just investment dollars - in capital

---

[24] Lenzner, *supra* note 22.

[25] For example, between 2008 and 2012, individual investors withdrew, on net, $500 billion from U.S. stock mutual funds, while adding more than $1 trillion to bond funds. Yousuf, *supra* note 12.

[26] Blake Ellis, *Savings Accounts with the Highest Yields*, CNN MONEY (Oct. 1, 2013), http://money.cnn.com/2013/10/01/pf/savings-account-yields.

[27] John Waggoner, *Bond Funds Give Investors a Lump of Coal in 2013*, USA TODAY (Dec. 30, 2013), http://www.usatoday.com/story/money/markets/2013/12/30/bond-funds-drop-in-2013/4240359.

[28] ROXBURGH, *supra* note 12, at 8.

[29] Zohar Goshen & Gideon Parchomovsky, *The Essential Role of Securities Regulation*, 55 DUKE L.J. 711, 714 (2006).

[30] *See* Ron Kaniel et al., *Individual Investor Trading and Stock Returns*, 63 J. OF FIN. 273 (2008).

markets is something to lament because of its effect on share price accuracy.

Historically, the primary goal of securities regulation has been viewed as protection of the "ordinary" investor.[31] Indeed, the SEC still considers this its most solemn charge.[32] Modern scholars of securities regulation, however, assert that the goal of securities regulation should be the attainment of efficient financial markets to improve the allocation of resources in the economy. For example, Zohar Goshen and Gideon Parchomovsky state:

> Any serious examination of the role and function of securities regulation must sidestep the widespread, yet misguided, belief that securities regulation aims at protecting the common investor. Securities regulation is not a consumer protection law. Rather, scholarly analysis of securities regulation must proceed on the assumption that the ultimate goal of securities regulation is to attain efficient financial markets and thereby improve the allocation of resources in the economy[33]... [T]his Article argues that information traders are the group that can best underwrite efficient and liquid capital markets, and, hence, it is this group securities regulation should strive to protect. By protecting information traders, securities regulation enhances efficiency and liquidity in financial markets.[34]

Of course, the goal of enhanced market efficiency 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.[35] However, Goshen and Parchomovsky define "information traders" as "sophisticated professional investors and analysts,"[36] not retail investors. Goshen and Parchomovsky's argument reflects the widely held belief that individual investors deserve no special protections and, in fact, are harmful to market functioning because they are "noise traders" that distort stock prices.[37]

Share prices are accurate when they reflect fundamental corporate value.

---

[31] Ralph K. Winter, *On 'Protecting the Ordinary Investor'*, 63 WASH. L REV. 881 (1988).

[32] *See* Mary Jo White, Chairwoman, Sec. & Exch. Comm'n, Protecting the Retail Investor, Address Before the Consumer Federation of America's Consumer Assembly (Mar. 21, 2014) ("[W]e are . . . focused on protecting the consumers in our securities markets - especially the individual investors, who we often refer to as "retail" investors - who invest their own money to save for retirement, or to buy a home or to send their children to college. The retail investor must be a constant focus of the SEC - if we fail to serve and safeguard the retail investor, we have not fulfilled our mission."), *available at* http://www.sec.gov/News/Speech/Detail/Speech/1370541226174.

[33] Goshen & Parchomovsky, *supra* note 31, at 713.

[34] *Id.* at 715 (citations omitted).

[35] Winter, *supra* note 31.

[36] Goshen & Parchomovsky, *supra* note 31, at 711.

[37] *See, e.g.*, Brad M. Barber et al., *Do Retail Trades Move Markets?*, 22 REV. FIN. STUD. 151 (2009).

They do this by incorporating information that predicts future cash flows to shareholders over the life of the firm.[38] 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. This noise trading-induced 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.[39] James Tobin describes the state of affairs when the stock market directs capital to its highest value use as "functional efficiency."[40] The term "allocative efficiency" is also used to describe this phenomenon. Functional or allocative efficiency requires accurate share prices.[41]

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 did not exist. Indeed, Donald Langevoort states:

> ...[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.[42]

Though publicly deeming individual investors "economic desirables" and eliminating them from the capital markets seems politically infeasible, it is not implausible. One leading securities regulation scholar, Stephen Choi, 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.[43] In addition, (1) the recent proliferation, and then consol-

[38] *See, e.g.*, Merritt B. Fox et al., *Law, Share Price Accuracy, and Economic Performance: The New Evidence*, 102 MICH. L. REV. 331 (2003) for a comprehensive discussion of the related concepts of "share price accuracy" and "share price informedness."

[39] Artyom Durnev et al., *Does Greater Firm-Specific Return Variation Mean More or Less Informed Stock Pricing?*, 41 J. ACCT. RES. 797 (2003).

[40] James Tobin, *On the Efficiency of the Financial System*, 153 LLOYD'S BANKING REV. 1 (1984).

[41] Durnev et al., *supra* note 39.

[42] Donald C. Langevoort, *Taming the Animal Spirits of the Stock Market: A Behavioral Approach to Securities Regulation*, 97 NW. U. L. REV. 135, 172-73 (2002).

[43] Stephen Choi, *Regulating Investors Not Issuers: A Market-Based Proposal*, 88 CALIF. L. REV. 279, 300 (2000).

idation, 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 that fully excludes individual investor participation is at least possible. Moreover, even if deliberately eliminating retail investors from the capital markets is seemingly infeasible, the presence of retail investors influences a great deal of SEC policy. Thus, understanding the market effect of retail investors has significant regulatory implications.

This Article builds on recent studies on the role of retail investors in markets and provides evidence that suggests that individual investors are not "economic undesirables" and that encouraging retail sector participation may improve market functioning. 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. Under this methodology, 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 higher levels of retail trading are associated with lower R^2's, but also demonstrates that retail investor trading levels are significantly correlated with PIN, a metric of price informativeness that measures the probability of informed trading in a particular stock. Moreover, the evidence suggests that there is reason to believe the relationships are causal (that is, retail trading causes changes in R^2 and PIN). The evidence, therefore, suggests that retail investors have a positive effect on share price accuracy and market efficiency, and therefore play an important role in market functioning.

The Article proceeds as follows. Part II summarizes the existing literature surrounding market efficiency and individual investors. Part III describes the two measures of stock price informativeness used in this study -- R^2 and PIN. Part IV describes the data used and analytical methodologies of this study. Part IV also presents preliminary results. Part V concludes.

II. Market Efficiency and Noise Trader Risk

Allocative efficiency requires capital to be directed to its highest and best use. Stock prices reflect not only past firm profitability, thereby rewarding companies for success, but also future opportunities for corporate value creation. If prices reflect informed judgments by market investors, stock prices are tools to help ensure that the companies that are the most profitable or efficient at providing desired goods and services receive the greatest share of investment capital. Accurate stock prices, therefore, are important for allocative efficiency.[44] The effect of irrational or noise traders on stock prices and market efficiency, a central issue in finance, therefore has important implications for the regulation of securities markets.

Historically, little attention was paid to the possibility of market risk from noise trading. The traditional belief was that irrational traders could not affect share prices over the long run. Under this theory, trading based on mistaken beliefs would lead to trading losses against rational, informed investors and wealth reductions that would make it impossible for irrational traders to survive in a competitive marketplace.[45] Trades of irrational traders were random and uncorrelated, thus tending to cancel one another out and largely eliminating any price effects from such trading.[46] Therefore, under this theory, there was little reason to worry about the presence of noise traders.

This traditional view has come under intense theoretical and empirical attack.[47] In a theoretical study, Kogan et al.[48] 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

[44] Of course, stock prices in and of themselves do not directly result in the allocation of capital in our society. James Dow & Gary Gorton, *Stock Market Efficiency and Economic Efficiency: Is There a Connection?*, 52 J. FIN. 1087, 1087 (1997). Stock prices, however, do serve a signaling function. First, stock prices transmit information from market investors to management about the value of future investment opportunities. *Id.* Second, stock prices transmit information to management about the company's past performance. Managers have discretion with respect to the level of investment and how to deploy the firm's resources. *Id.* However, incentives based on informed stock prices, through for example, stock options, guide managerial choices to optimal investment. *Id.*

[45] *See, e.g.*, MILTON FRIEDMAN, *The Case for Flexible Exchange Rates*, *in* ESSAYS IN POSITIVE ECONOMICS 157 (1953); Eugene F. Fama, *The Behavior of Stock Market Prices*, 38 J. BUS. 34 (1965).

[46] *See, e.g.*, FRIEDMAN, *supra* note 45; Fama, *supra* note 45.

[47] As noted in Leonid Kogan et al., *The Price Impact and Survival of Irrational Traders*, 61 J. FIN. 195, 195-196 (2006) and J. Bradford De Long et al., *Noise Trader Risk in Financial Markets*, 98 J. POL. ECON. 703 (1990), irrational traders hold portfolios with high growth and can potentially outgrow rational traders and thus survive on the basis of a partial equilibrium model. *But see* Alvaro Sandroni, *Do Markets Favor Agents Able to Make Accurate Predictions?*, 68 ECONOMETRICA 1303 (2000); Lawrence Blume & David Easley, *If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets*, 74 ECONOMETRICA 929 (2006) (using general equilibrium models to conclude that "irrational traders do not survive in the long run," Kogan et al., *supra* at 196).

[48] Kogan et al., *supra* note 47.

another out. This evidence suggests that noise traders often act as a herd.[49] 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.[50]

Most prior research shows that individual investors are more likely to make irrational or imprudent investment decisions than institutional investors,[51] and hence, are the primary suspects in the search for noise traders. Brad Barber and Terrance Odean plainly state, "with some notable exceptions,...the evidence indicates that individual investors are subpar investors." [52] For example, Odean,[53] after analyzing the trading records of 10,000 retail investors at a large discount brokerage firm, finds that stocks *purchased* by retail investors underperformed stocks *sold* by retail investors by 23 basis points (0.23%) per month over the following year. This result holds even after excluding trades that likely were not driven by fundamentals, but rather were in response to liquidity, rebalancing or tax needs. The statistical significance of the result is not strong (p-value of approximately 0.07), but Odean concludes that retail investors have "perverse security selection ability."[54] Grinblatt and Keloharju,[55] after analyzing two years of investment return data in Finland, conclude that retail investors are "net buyers of stocks with weak future performance."[56] Barber et al.,[57] using Taiwanese trading records from 1995-1999, construct investment portfolios that mimic institutional and retail investor behavior. Barber et al.[58] find that the investment strategy that mimics retail investor trading over a 140-trading day period earns a negative return of 75 basis points (0.75%) per month before factoring in transaction costs.[59]

---

[49] For a review of the literature related to herding behavior, see David Hirshleifer & Siew Hong Teoh. *Herd Behaviour and Cascading in Capital Markets: A Review and Synthesis*, 9 EUR. FIN. MGMT. 25 (2003); Thomas Lux, *Herd Behaviour, Bubbles and Crashes*, 105 ECON. J. 881 (1995).

[50] *See, e.g.*, Goshen & Parchomovsky, *supra* note 31; Langevoort, *supra* note 42.

[51] Andrew Jackson, The Aggregate Behaviour of Individual Investors (2003) (unpublished manuscript), *available at* http://ssrn.com/abstract=536942.

[52] Brad M. Barber & Terrance Odean, *The Behavior of Individual Investors*, *in* HANDBOOK OF THE ECONOMICS OF FINANCE 1533, 1539 (George M. Constantinides et al. eds., 2013). Barber and Odean note that a lot of the reason for the poor performance of retail investors is costs (e.g., commissions), but also point out that retail investors incur trading losses before factoring in transaction costs. *Id.*

[53] Terrance Odean, *Do Investors Trade Too Much?*, 89 AMER. ECON. REV. 1279 (1999),.

[54] Barber & Odean, *supra* note 52, at 1540.

[55] Mark Grinblatt & Matti Keloharju, *The Investment Behavior and Performance of Various Investor Types: A Study of Finland's Unique Data Set*, 55 J. FIN. ECON. 43 (2000).

[56] Barber & Odean, *supra* note 52, at 1541.

[57] Brad M. Barber et al., *Just How Much Do Individual Investors Lose by Trading?*, 22 REV. FIN. STUD. 609 (2009).

[58] Barber et al., *supra* note 57.

[59] Barber & Odean, *supra* note 52, at 1541.

Not only is there evidence of poor performance by retail investors, but there is also evidence their poor skills can have market effects. Barber, Odean and Zhu[60] find that (1) small trades (serving as a proxy for trades by individual investors) are correlated, (2) buying by retail investors pushes prices too high (above their fundamental values), and (3) selling by retail investors pushes prices too low (below their fundamental values). They conclude that individual investors (whom they term noise traders) can move equity markets. Similarly, Hvidkjaer,[61] 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[62] also find that (1) retail investor trades are systematically correlated, (2) this concerted action can affect stock returns, and (3) this "retail sentiment" does not appear to be an outgrowth of a reaction to factors related to fundamental value.

Not all commentators view noise traders as having purely negative effects on market efficiency. One argument is that noise trading 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.[63] Thus, noise trading makes markets more liquid as informed traders attempt to exploit inefficiencies in markets caused by noise trading.[64] 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.[65] Unfortunately for the noise traders, the way noise traders make this worthwhile for informed traders is by suffering losses in trades against informed traders.[66]

Despite the prevalent view that retail investors are noise traders, there is recent evidence that suggests that individual investor trades predict future perfor-

---

[60] Barber et al., *supra* note 37.

[61] Soeren Hvidkjaer, *Small Trades and the Cross-Section of Stock Returns*, 21 REV. FIN. STUD. 1123 (2008).

[62] Alok Kumar & Charles M. C. Lee, *Retail Investor Sentiment and Return Comovements*, 61 J. FIN. 2451 (2006).

[63] *See* Fischer Black, *Noise*, 41 J. FIN. 529 (1986).

[64] *Id.* 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., *supra* note 47, at 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.").

[65] Gregory La Blanc & Jeffrey J. Rachlinski, *In Praise of Investor Irrationality*, *in* LAW AND ECONOMICS OF IRRATIONAL BEHAVIOR 542 (Francesco Parisi & Vernon L. Smith eds., 2005).

[66] *Id.*

mance and, therefore, that retail investors are not noise traders, but rather *informed* investors. Barber and Odean note that there is intriguing evidence that stocks *purchased* by individual investors over a short time horizon (e.g., a day or a week) go on to earn superior returns in the next week, while stocks *sold* by individual investors go on to earn poor returns.[67] For example, Kaniel, Saar, and Titman,[68] using data on individual investor trades from the New York Stock Exchange's Consolidated Audit Trail Data (CAUD) files over the period 2000-2003, find that the top decile of stocks purchased by retail investors earned market-adjusted returns of 16 basis points (0.16%), while the bottom decile of stocks sold by retail investors earned -33 basis points (-0.33%) over the subsequent 20 trading days. The authors conclude that the results are consistent with retail investors serving as liquidity sources for institutional investors who require immediacy. However, using the same data, Kaniel et al.,[69] find that stocks purchased by retail investors in the 10 days before corporate earnings announcements outperformed stocks sold by individual investors by 1.5% in the two days around the earnings announcement. The authors conclude that informed trading by individual investors is at least as responsible for this result as is the provision of liquidity to institutional investors. Another study conducted by Barber, Odean, and Zhu shows that stocks heavily bought by individuals (as proxied for by small trade size) outperform stocks heavily sold by individuals in the subsequent two weeks (though they then go on to underperform for the rest of the year).[70] Finally, Kelley and Tetlock,[71] using retail brokerage data for the period 2003-2007, find that retail trades positively predict stock returns up to 20 days post-trade. The researchers conclude "retail market orders aggregate private information about firms' future cash flows . . ."[72] As Barber and Odean note, the debate surrounding the origins of this documented U.S. retail investor savvy in the short-run continues,[73] but the foregoing studies provide strong evidence that retail investors, as a group, are able to positively predict future returns.[74]

---

[67] Barber & Odean, *supra* note 52, at 1539.

[68] Ron Kaniel et al., *Individual Investor Trading and Stock Returns*, 63 J. FIN. 273 (2008). .

[69] Ron Kaniel et al., *Individual Investor Trading and Return Patterns Around Earnings Announcements*, 67 J. FIN. 639 (2012),.

[70] Barber et al., *supra* note 37 (explaining the superior short-run performance followed by negative long-run performance as resulting from "the correlated sentiment-based trading of individual investors. In the short run, sentiment temporarily pushes prices above fundamental value, leading to predictable long-run return reversals"); Barber & Odean, *supra* note 52, at 1543.

[71] Eric K. Kelley & Paul C. Tetlock, *How Wise Are Crowds? Insights from Retail Orders and Stock Returns*, 68 J. FIN. 1229 (2013).

[72] Barber & Odean, *supra* note 52, at 1543.

[73] *Id.*

[74] Barber and Odean reconcile this evidence of short-run return predictability with the general evidence that individual investors fare poorly in investing by noting that though

The evidence on the investment skill of non-US retail investors is mixed,[75] but the following studies of international markets suggest that retail investors are not noise traders and may even aid market efficiency. First, Henker and Henker examine a sample of small capitalization stocks trading on the Australian stock market that experience an asset price bubble between December 1, 1995, and March 1, 2002.[76] The researchers conclude that, even in an environment where retail investors are likely to have the most influence (i.e., trading in small cap stocks with a 55% retail trading participation rate in the sample firms),[77] individual investors were not responsible for these stock mispricings.[78]

Second, Henker and Paul, in a study of retail trades in Australian small capitalization stocks, reject the view that retail investors are the cause of the "January" effect (i.e., the large anomalous returns earned for stocks in January that cast doubt on market efficiency).[79] The researchers do note the presence of a January effect in their sample of Australian stocks between December 1995 and December 2001, but conclude that retail trading is not responsible for the anomaly.[80] They further "challenge the theory that retail investors, as irrational noise traders, are responsible for market anomalies."[81] That said, Henker and Paul, consistent with Henker and Henker, do conclude that retail trading "has a negligible impact on market prices."[82]

In addition, Jackson,[83] after analyzing a unique dataset of 41.9 million retail investor trades over an eleven-year period on the Australian Stock Exchange,[84] 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 individu-

---

retail investors have been shown to perform well over the short-term, they, on average, hold stocks for 16 months. Thus, any short-term gains are easily eclipsed by longer-term losses. *Id.* at 1539.

[75] *Id.* at 1543.

[76] Julia Henker & Thomas Henker, *Are Retail Investors the Culprits? Evidence from Australian Individual Stock Price Bubbles*, 16 EUR. J. FIN. 281, 289 (2010).

[77] *Id.* at 289.

[78] *Id.* at 281.

[79] Julia Henker & Debapriya J. Paul, *Retail Investors Exonerated: The Case of the January Effect*, 52 ACCT. & FIN. 1083 (2012).

[80] *Id.* at 1097.

[81] *Id.* at 1098.

[82] *Id.*

[83] Jackson, *supra* note 54.

[84] According to Jackson, as of 2001, the U.S. market was 37 times as large as the Australian market (as measured by total market capitalization). Also, in the U.S. retail investors owned approximately 42% of the stock traded on U.S. markets. In Australia, the percentage of individual investor ownership at that time was 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. *Id.*

als positively predict future market returns.[85] Jackson states that one potential reason for this result could be individuals' possession of valuable private information.[86]

Finally, Choe, Kho, and Stulz,[87] based on a study of two years (1996-1998)[88] of Korea Stock Exchange trading data, find that domestic retail investors[89] 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 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.[90]

III. Measures of Stock Price Informativeness

This Article contributes to the debate on the market effects of retail trading by examining the relationship between retail trading activity, on the one hand, and two measures of stock price informativeness— R^2 and PIN, on the other.

A. R2

In this study, the first measure of stock price informativeness I use is R^2. "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

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

[86] *Id.*

[87] Hyuk Choe et al., *Do Domestic Investors Have More Valuable Information About Individual Stocks than Foreign Investors?* (Nat'l Bureau Econ. Research, Working Paper No. 8073, 2001), *available at* http://www.nber.org/papers/w8073.

[88] 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.

[89] Choe, Kho, and Stulz 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. Choe et al., *supra* note 87.

[90] The database on which Choe, Kho, and Stulz 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. *Id.*

return) is explained[91] 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 information is noise, rather than information related to a company's fundamentals.[92]

Available evidence that a low R^2 is a measure of informational efficiency and share price accuracy is strong. Durnev et al.[93] 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.[94] 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[95] find that firms operating in U.S. industries with lower R^2s use more external financing. The authors suggest that this

---

[91] The word "explained" as used in this context does not suggest that the independent variables *cause* changes in the dependent variable.

[92] Richard Roll, *R^2*, 43 J. FIN. 541 (1988); Durnev et al., *supra* note 39. 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 Part and Appendix A, researchers have provided evidence on whether one or the other explanation is more likely, rather than a little of both.

[93] Durnev et al., *supra* note 39.

[94] *Id.*

[95] Artyom Durnev et al., *Does Firm-Specific Information in Stock Prices Guide Capital Allocation?* (Nat'l Bureau Econ. Research Working Paper No. 8093, 2001).

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[96] find a strong correlation between firm-specific return variation and 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 suggest,[97] may use this signal to intervene as necessary in the face of poor management decisions. Similarly, Chen, Goldstein, and Jiang[98] present evidence that fluctuations in stock prices affect the capital investment decisions of firms with low R^2s more than those with high R^2s. 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[99] calculate, inter alia, the average R^2s 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^2s) 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^2s. 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[100] and less country-level opaqueness.[101]

---

[96] Artyom Durnev et al., *Value-Enhancing Capital Budgeting and Firm-specific Stock Return Variation*, 59 J. FIN. 65 (2004).

[97] *Id.*

[98] Qi Chen et al., *Price Informativeness and Investment Sensitivity to Stock Price*, 20 REV. FIN. STUD. 619 (2007).

[99] Randall Morck et al., *The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements?*, 58 J. OF FIN. ECON. 215 (2000).

[100] Jeffrey Wurgler, *Financial Markets and the Allocation of Capital*, 58 J. OF FIN. ECON. 187 (2000).

[101] Li Jin & Stewart C. Myers, *R^2 Around the World: New Theory and New Tests*, 79 J. FIN. ECON. 257 (2006). Jin and Myers define opaqueness as a "lack of information that would enable investors to observe operating cash flow and income and determine firm value." *Id.* at

The R^2 methodology also has found adherents among the ranks of legal scholars. For example, Fox, Morck, Yeung, and Durnev[102] conclude that enhanced mandatory disclosure rules adopted in the United States in December 1980 made share prices more accurate (as evidenced by a decrease in average R^2 across firms). Beny[103] employs the country R^2 statistics of Morck, Yeung, and Yu[104] 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, some scholars question the informational-efficiency interpretation of R^2 and contend instead that R^2 is a measure of price *inefficiency*. Other scholars have entered the debate by attempting to reconcile these two competing interpretations of R^2.[105] Though the weight of the evidence, in my view, supports the conclusion that low R^2 is a sign of informational efficiency, particularly in the context of firms trading in informationally efficient environments such as that of the NYSE companies examined in this study, I also employ a second measure of price informedness as a robustness check.

B. PIN

PIN, a widely-used[106] measure of private information based on the microstructure model developed by Easley, Kiefer, and O'Hara (the "EKO Model"),[107] estimates the probability of informed trading in a particular stock over a specified time period using observed order flow.[108] The model makes three key assumptions:[109]

> 1. In the presence of "common (i.e., easily understandable and public) knowledge," a stock's specialist or market maker will move prices to appropriate levels automatically without any trading activity.

281.

[102] Fox et al., *Law, Share Price Accuracy, and Economic Performance: The New Evidence*, 102 MICH. L. REV. 331 (2003).

[103] Laura Nyantung Beny, *Do Insider Trading Laws Matter? Some Preliminary Comparative Evidence*, 7 AM. L. & ECON. REV. 144 (2005).

[104] Morck et al., *supra* note 99.

[105] See Appendix A for a discussion of the debate.

[106] *See e.g.*, Stephen Brown & Stephen A. Hillegeist, *How Disclosure Quality Affects the Level of Information Asymmetry*, 12 REV. ACCT. STUD. 443 (2007); Clara Vega, *Stock Price Reaction to Public and Private Information*, 82 J. FIN. ECON. 103 (2006); Chen et al., *supra* note 98.

[107] David Easley et al., *One Day in the Life of a Very Common Stock*, 10 REV. FIN. STUD. 805 (1997).

[108] Brown & Hillegeist, *supra* note 106, at 448.

[109] *See* Vega, *supra* note 106, at 106.

2. A stock's order flow reflects trading on information that is not "common knowledge." Neither (a) private information nor (b) public information that certain investors are particularly skillful at analyzing is considered "common knowledge."[110] Having either of these types of information leads traders to buy or sell. Thus, "abnormal order flow," defined as excess buying or selling pressure on a stock, captures information that is not common knowledge. This is "informed trading."

3. If trading is not "informed trading," it is "noise trading,"[111] which includes not only trading based on irrational beliefs, but also trading for liquidity-based reasons that are unrelated to information about the stock in question.

Vega describes the PIN model and calculation of PIN as follows:

> The game consists of three players, liquidity [or noise] traders, informed traders and a market maker... Liquidity traders buy or sell shares of the asset for reasons that are exogenous to the model[,] and each buy and sell order arrives to the market according to an independent Poisson distribution with a daily arrival rate equal to ε. The probability that an information event occurs is a, in which case the probability of bad news is σ and the probability of good news is $(1-\delta)$. If an information event occurs, the arrival rate of informed traders is μ. Informed traders trade for speculative reasons; if they receive good news (the current asset price is below the liquidation value of the asset)[,] they buy one share of the asset[.] [I]f they receive bad news[,] they sell one share of the asset.
>
> On days with no information events, which occur with probability $(1-a)$, the arrival rate of buy orders is ε[,] and the arrival rate of sell orders is ε as well. Thus, the total amount of transactions on noninformation days is 2ε with the number of buys approximately equal to the number of sells. On a bad information event day, which occurs with probability $a\delta$, we observe more sells than buys. To be precise, the arrival rate of buy orders is ε and the arrival rate of sell orders is $\varepsilon+\mu$. In contrast, on a good information event day, which occurs with probability $a(1-a)$, we observe more buys than sells, i.e., the arrival rate of buy orders is $\varepsilon+\mu$ and the arrival rate of sell orders is ε.
>
> [To describe this mathematically,] ...PIN [is] the estimated arrival rate of informed trades divided by the estimated arrival rate of all trades during a pre-specified period of time. Formally,
>
> PIN = $\underline{a\mu}$

[110] *Id.* at 105.
[111] Chen et al., *supra* note 98, at 627.

$\alpha\mu + 2\varepsilon$ (Internal citations omitted)[112]

Put another way, to calculate PIN, one uses secondary market trade data to estimate normal and abnormal trading, and, with these estimates, one calculates the ratio of abnormal trading to total trading.[113] The PIN methodology assumes informed traders will cluster their trades following new information (e.g., buy orders will be clustered following good news events and sell orders will be clustered following bad news events) and that uninformed trades will be unclustered. Since abnormal trading is assumed to represent informed trading,[114] the ratio of abnormal trading to total trading becomes the ratio of informed trading to total trading. The higher the PIN, the higher the level of informed trading in a particular stock.

Though the basic EKO model is widely used in the finance literature, because the model assumes that uninformed buy and sell orders are uncorrelated (or unclustered) (which is often not the case in the real world), some researchers extend the model to adjust for correlation among, and trading intensity of, uninformed buyers and sellers.[115] PIN, thus, is calculated under this extended model, where $v = \mu/\varepsilon$, as:[116]

$$\text{PIN} = \underline{\alpha\mu} = \underline{\alpha\varepsilon v} = \underline{\alpha v}$$

---

[112] Vega, *supra* note 106, at 106-07.

[113] Ozgur (Ozzy) Akay et al., *What Does PIN Identify? Evidence from the T-Bill Market*, 15 J. FIN. MKTS. 29, 30 (2012).

[114] *Id.* (arguing that in testing the theory set forth by Duarte and Young, Jefferson Duarte & Lance A. Young, *Why is PIN Priced?* 91 J. FIN. ECON. 119 (2009), abnormal order flow imbalances may represent not only informed trading, but that clusters also could represent "liquidity shocks or the effect of the changes in the demand for immediacy"). Thus, Akay et al. question the use of PIN as a measure of informed trading. Aslan et al., responding to Duarte and Young (and, by implication, Akay et al.), argue: "[Duarte and Young] view this liquidity effect as being unrelated to information and thus argue that [it is] liquidity and not information that is being priced. But this leaves unanswered the question of the source of the common shocks. To the extent that the shocks are the result of private interpretations of public information they, too, reflect information-based trade. Our results here showing that PIN is correlated with a variety of variables widely acknowledged in the literature as information proxies strongly suggests that it does relate to information." Hadiye Aslan et al., *The Characteristics of Informed Trading: Implications for Asset Pricing*, 18 J. EMPIRICAL FIN. 782, 784 (2011).

[115] As Brown and Hillegeist explain, "[a]n important assumption of the [basic EKO] model is that the daily arrival rates of uninformed buy and sell orders are drawn from independent Poisson distributions with constant parameters; as such, the daily numbers of uninformed buys and sells are uncorrelated. However, in practice, private information events (such as the release of macroeconomic statistics and earnings announcements) often affect the trading intensity of all uninformed traders - both buyers and sellers - on a particular day so that the daily arrival rates of uninformed buy and sell orders are positively correlated . . . . To relax this restrictive assumption, Venter and de Jongh...model the arrival of uninformed buy and sell orders...[and cause] the average trading intensities of uniformed investors, both buyers and sellers, . . . [to be] subject to a daily scaling factor...Hence, the extended model allows for a positive correlation between the daily number of buy and sells . . . ." Brown & Hillegeist, *supra* note 106, at 448-449.

[116] *Id.* at 449.

$$a\mu + 2\varepsilon \qquad a\varepsilon v + 2\varepsilon \qquad av + 2\varepsilon$$

This equation demonstrates that the PIN (probability of informed trading) is higher with (1) increased frequency of private information events (*a*), (2) increases in absolute and relative informed investor trading intensity (μ and v), and (3) decreases in uninformed investor trading intensity (ε).[117] I use this extension of the EKO Model as the PIN variable in this study.

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

A. Data and Sample

The period of this study is April 1, 2005 - August 31, 2006. I obtained data on firm-level retail trading activity, including total shares purchased and sold by retail investors on the NYSE each day[118] for a particular stock[119] from the NYSE ReTracEOD Summary.[120] 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

---

[117] *Id.*

[118] 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.

[119] 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 the NYSE. *See, e.g.*, Michael Goldstein et al., Competition and Consolidation in the Market for NYSE-Listed Securities (2006) (unpublished manuscript), *available at* http://www.fma.org/SLC/Papers/Competition_and_Consolidation_in_the_Market_for_the_NYSE-listed_securities.pdf. 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.

[120] 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, NYSE ORDER TRACKING SYSTEM, INPUT FILE LAYOUT, v. 4.3 (2004), *available at* http://www.nyse.com/pdfs/order_tracking_system_v4.3b.pdf. 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.

such as small share size[121] or odd-lot trading.[122] Those proxies suffer from some important limitations. 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.[123] 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.[124] 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 to do so.[125] Use of odd-lot data loses the impact of round-lot trading in study results. Finally, using small trades as a proxy for individual investor trading post-2001 is problematic because of (1) the adoption of price decimalization in 2001 and (2) the more prevalent use of algorithmic (computerized) trading.[126] Both developments led to substantial increases in the volume of small trades executed by institutional investors.[127]

The study employs quarterly PIN data used in Brown and Hillegeist.[128] 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

---

[121] Barber, Odean and Zhu recognize the limitations of small share size as a proxy for individual investor trading and do some limited testing of the data used in their study against actual brokerage firm data to gain comfort in the representativeness of their data sets. Barber et al., *supra* note 70.

[122] Odd-lot trading is trading a number of shares other than that which is required for a round lot (100 shares). Wu, 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. Hsiu-Kwang Wu, *Odd-Lot Trading in the Stock Market and Its Market Impact*, 7 J. FIN. & QUANTITATIVE ANALYSIS 1321 (1972). For an example of a more recent work using odd-lot trading data as a proxy for retail investor trading, see Jorge Brusa et al., *Weekend Effect, 'Reverse' Weekend Effect, and Investor Trading Activities*, 32 J. BUS., FIN. & ACCT, 1495 (2005).

[123] Hvidkjaer, *supra* note 61.

[124] Indeed, Patterson and Lucchetti note that the average size of a market trade as of 2008 was 260 shares, down from 1,400 shares a decade prior. Scott Patterson & Aaron Lucchetti, *Boom in "Dark Pool" Trading Networks Is Causing Headaches on Wall Street*, WALL ST. J., May 8, 2008.

[125] See note 166 *infra* for further discussion.

[126] Barber & Odean, *supra* note 52, at 1541 n.3.

[127] *Id.*

[128] Brown & Hillegeist, *supra* note 106. I am grateful to Professor Stephen Brown for making these data available on his website at http://www.rhsmith.umd.edu/faculty/sbrown/.

from the merged CRSP-Compustat database.[129] Institutional ownership data, based on the 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 for the R^2 and PIN analyses, I began with every NYSE-listed common stock[130] in the CRSP database during the study period.[131] 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.[132] My final sample consists of 1,129 different stocks for the R^2 analysis and of 1,126 for the PIN analysis. 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) for the R^2 analysis and into two equal groups of 563 different stocks for the PIN analysis. I refer to these groups as "Top Half" (the larger firms) and "Bottom Half" (the smaller firms).

B. Firm-Specific Stock Return Variation (R2)

As the first measure of share price informedness or accuracy, I use firm-specific stock return variation. Following the model of Durnev et al.[133] and others, I obtain firm-specific stock return variation (R^2) by use of the following regression:

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

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

[129] See Part IV.D. for a description of the accounting data used in this study.

[130] 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).

[131] 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.

[132] 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.

[133] Durnev et al., *supra* note 39.

(defined as all firms in the same two-digit SIC code), excluding the firm in question.[134] 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.

C. PIN

As the second measure of stock price informedness, I use PIN. For PIN, I use the average of the quarterly PINs calculated by Brown and Hillegeist[135] for each quarter from April 2005 to September 2006.[136]

Empirical Methodology and Results of Analysis

(i) R2

My objective is, first, 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[137] of R^2, New R^2, as my dependent variable.[138] I obtain New R^2 by using the following formula:

$$New\ R^2 = ln(R^2_i/1 - R^2_i) \qquad (2)$$

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

[134] Consistent with Durnev et al., *id.*, 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.

[135] Brown & Hillegeist, *supra* note 106.

[136] Because the study period ends on August 31, 2006, I adjust the third quarter 2006 PIN in the mean calculation.

[137] This is a common econometric remedy. *See, e.g.*, Morck et. al, *supra* note 99. The transformed variable is a continuous variable that is more normally distributed than R^2, which has values between 0 and 1. *See* Hollis Ashbaugh-Skaife et al., *Does Stock Price Synchronicity Represent Firm-Specific Information? The International Evidence* (MIT Sloan, Research Paper No. 4551-05, 2006), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=768024.

[138] In unreported results, I run the regressions described 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 this Part.

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 ("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 stock price informativeness. I control for level of "INSTITUTIONAL OWNERSHIP" (defined as the proportion of a firm's stock held by institutions as of March 31, 2005)[139] 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,[140] 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 indi-

[139] 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, e.g.*, ANJAN V.THAKOR. JEFFREY S. NIELSEN, & DAVID A. GULLEY, THE ECONOMIC REALITY OF SECURITIES CLASS ACTION LITIGATION (2005). Thus, the figures, though anomalous, should not bias this study's results significantly. In the instant study, sixty-five of the 1,129 stocks in the sample (5.8%) have institutional ownership percentages that, as calculated, exceed 100%.

[140] *See, e.g.*, Veronica Pizarro et al., The Influence of Insiders and Institutional Owners on the Value, Transparency, and Earnings Quality of Chilean Listed Firms? (Apr. 27, 2007) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=982697.

viduals or institutions with 5% or greater stock ownership in the company ("5% OWNER OWNERSHIP").[141] 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.[142] In addition, I control for firm news coverage ("NEWS COVERAGE") during the study period[143] 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[144] 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

[141] Perhaps 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.

[142] 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. P. "Rajan" Varadarajan & Vasudevan Ramanujam, *Diversification and Performance: A Reexamination Using a New Two-Dimensional Conceptualization of Diversity in Firms*, 30 ACAD. MGMT. J. 380 (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.

[143] 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 Part.

[144] Brad M. Barber & Terrance Odean, *All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors*, 21 REV. FIN. STUD. 785 (2008).

Baker and Wurgler[145] 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 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.[146] 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);[147] two variables related to profitability - "RETURN ON EQUITY"[148] and a dummy variable for whether a firm is profita-

---

[145] Malcolm Baker & Jeffrey Wurgler, *Investor Sentiment and the Cross-Section of Stock Returns*, 61 J. FIN. 1645 (2006).

[146] As Baker and Wurgler 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." *Id.* at 1649-50 (internal citations omitted).

[147] 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 Part 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.

[148] 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

ble (that is, has positive earnings); two variables related to dividend payments - "DIVIDENDS-TO-EQUITY"[149] 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;"[150] and three variables to proxy for characteristics indicating high growth opportunities and/or distress - "BOOK-TO-MARKET EQUITY,"[151] "LEVEL OF EXTERNAL FINANCE,"[152] and "SALES GROWTH."[153] 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[154] include buy-side retail trading, sell-side re-

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).

[149] 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 148 above.

[150] 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, *supra* note 145, missing values of R&D expense are set to zero. As a robustness check, I run the regression analysis described in this Part 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.

[151] Book-to-market equity is book equity (defined in note 148 above) divided by a firm's average market capitalization during the study period.

[152] 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).

[153] Sales growth is the change in net sales (Compustat Item 12) from 2003 to 2004 divided by 2003 net sales. For this analysis, in addition to this variable for sales growth, I create sales growth dummies for the following categories of firms. 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."

[154] 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 Part. I, however,

tail trading, market capitalization, trading volume, research activity, insider ownership, dividends-to-equity, R&D-to-assets and age in months.[155] As a check against outliers and influential data points,[156] I winsorize all independent variables at their 0.5% and 99.5% values, as in Baker and Wurgler.[157]

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.[158] 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[159] because they are highly correlated with the sales growth variable.

My regression takes the form:

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

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

---

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. Thus, a logarithmic transformation is impossible for these variables.

155 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.

156 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.

157 Baker & Wurgler, *supra* note 145.

158 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.

159 *See* note 153.

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 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).[160]

(ii) PIN

My second objective to is examine the relationship between PIN and retail trading activity. I, therefore, substitute PIN as the dependent variable in the regression described above. My regression, therefore, takes the form:

$$PIN = \alpha + \beta RETTRADE_i + \gamma_1 I_i + \gamma_2 SIZEVOLRES_i + \gamma_3 INSOWN_i + \gamma_4 5PEROWN_i + \gamma_5 DIVERSE_i + \gamma_6 NEWSDAYS_i + \gamma_7 AGE_i + \gamma_8 STKVOL_i + \gamma_9 RETONEQ_i + \gamma_{10} DIVTOEQ_i + \gamma_{11} TANGASSETS_i + \gamma_{12} RDTOASSETS_i + \gamma_{13} BKTOMKT_i + \gamma_{14} EXTFIN_i + \gamma_{15} SALESG_i + \varepsilon_i \quad (4)$$

The control variable definitions are identical to those described previously in the regression that uses R^2 as the dependent variable. Table 3 reports the results of this regression, including t-statistics and robust standard errors, for the overall sam-

[160] Some firms move more closely with the market because they are more sensitive to general economic conditions. See Durnev et al., *supra* note 39, for a general discussion of this point. Thus, it would not be fair to say that the stock prices of such firms (i.e., firms with high R^2's) 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. I do not repeat this analysis with PIN (described below) as the dependent variable because there are not similar concerns about the relationship between macroeconomic factors and the probability of informed trading.

ple. These results show that higher proportions of retail trading activity are positively associated with PIN. This result, too, is statistically significant (p-value < 0.01). These results, combined with the R^2 regression results, suggest that increased levels of retail trading are associated with more informed share prices.

(iii) Size Effects

I also examine the extent to which firm size matters with respect to R^2 and PIN. 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. This finding implies that larger NYSE firms, as a group, have less accurate stock prices than smaller firms. However, recall 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,[161] 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 larger half of the firms in the sample (the "Top Half") and then only the smaller half of the 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[162] firms.

Table 1 also reveals that, on average, the largest firms in the sample (the "Top Half") have lower PIN's than the smallest firms in the sample (the "Bottom Half"). This is consistent with findings associated with R^2 as described above. This is suggestive of the fact that there is, in general, less private or difficult to interpret public information with respect to large firms prior to the study period.

As was performed with R^2 as described above, I perform separate regressions, using PIN as the dependent variable, including, first, only the larger half of the firms in the sample (the "Top Half") and then only the smaller half of the firms in the

---

[161] Sudipto Dasgupta et al., *Transparency, Price Informativeness, Stock Return Synchronicity: Theory and Evidence*, 45 J. OF FIN. & QUANTITATIVE ANALYSIS 1189 (2010). See Appendix A for a description of Dasgupta, Gan and Gao's findings.

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

sample (the "Bottom Half"). The results in Table 3 reveal that higher levels of retail trading are associated with higher PIN's in the overall sample and in the Top Half and Bottom Half size groups. However, unlike in the R^2 regression, there is a statistically significant relationship between retail trading activity and PIN in not only the overall sample and the Bottom Half size group, but also in the Top Half size group.

(iv) Causation

Though the results described above demonstrate that retail investor trading is correlated with firm-specific return variation (R^2) and probability of informed trade (PIN), I have not established that retail investor trading *causes* changes in R^2 or PIN. There may be no causal link or the causation may run in the opposite direction. For example, with respect to R^2, 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 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 more publicity and, thus, attract more retail investors. Consistent with that hypothesis, Barber and Odean[163] 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.[164] On the other hand, with respect to PIN, it is hard to imagine a scenario in which retail investors would be attracted to firms with high levels of private information-based trades, thus making reverse causation unlikely, but it still is not possible, from the analysis thus far, to determine if the relationship between retail trading and PIN is a causal one. 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 or PIN, but instrumental variable (IV) estimation is widely used by econometricians to determine the existence of causal relationships and address simultaneity concerns. In this study, I conduct two separate analyses. Through the two-stage least squares method of IV estimation, I, in the first stage, predict retail trading by

---

[163] Barber & Odean, *supra* note 144.

[164] 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. The 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.

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. I then repeat this analysis for PIN. "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 prices in the IV estimation are winsorized at their 0.5% and 99.5% values.

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 (or PIN), 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[165] 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 who prefers to purchase stock in round lots.[166] To this investor, a stock with a price of $9 may be significantly more attractive than one trading at $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.[167] Grullon, Kanatas, and Weston[168] 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.[169]

---

[165] Gustavo Grullon et al., *Advertising, Breadth of Ownership, and Liquidity*, 17 REV. FIN. STUD. 439 (2004).

[166] The reason for a round lot preference likely is not rooted in concerns about transaction cost differentials. Angel 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 reluctant to trade in odd lots." James J. Angel, *Picking Your Tick: Toward a New Theory of Stock Splits*, 10 J. APPLIED CORP. FIN. 59, 62 (1997). Similarly, Dhar et al., who 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. Ravi Dhar et al., The Impact of Clientele Changes: Evidence from Stock Splits 19 (2004) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=410104.

[167] *See* Dhar et al., *supra* note 166.

[168] Grullon et al., *supra* note 165.

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

The results of the IV analysis using R^2 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)[170] 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 one of the main regressions in this study (R^2) is derived from a regression whose dependent variable is stock price 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 in the case of R^2 because there is some evidence that returns can be related to absolute stock price level. For example, Gaunt, Gray and McIvor, in a study on Australian equity market returns, find that share prices, independent of firm size, affect portfolio returns.[171] Similarly, Bhardwaj and Brooks find a low price stock January effect (that is, the stocks earn abnormal returns in January).[172] Bhardwaj and Brooks characterize this

[170] Pairwise correlation coefficient with significance at the 1% level.

[171] Clive Gaunt et al., *The Impact of Share Price on Seasonality and Size Anomalies in Australian Equity Returns*, 40 ACCT. & FIN. 33, 33 (2000).

[172] Ravinder K. Bhardwaj & Leroy D. Brooks, *The January Anomaly: Effects of Low Share Price, Transaction Costs, and Bid-Ask Bias*, 47 J. FIN. 553, 553 (1992). Note that these abnormal

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 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.[173] 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.[174] Therefore, the available evidence suggests that stock price is a valid instrument for retail trading in this context.[175]

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.[176]

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

---

returns disappear when factoring in transaction costs and bid-ask bias to returns in the 1977-1986 period.

[173] *Id.* at 559.

[174] Bhardwaj and Brooks 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. *Id. at 556.* 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 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.

[175] 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 174 above, few of the firms in this study's sample are low-priced stocks.

[176] 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 Part. 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.

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 (R^2).

The results of the IV analysis using PIN for the overall sample appear in Table 5. Consistent with the findings with respect to R^2 above, 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 = 65.72). In addition, as stated previously, the correlation between retail trading and stock price (-0.45) is strong. Many of the qualitative reasons for believing stock price is a valid instrument that apply with respect to R^2 also apply with respect to PIN. Absolute stock price level should bear no weight on the likelihood of private information being impounded into stock prices or differing interpretations of public news.

The second stage regression demonstrates the effect retail trading has on PIN. This regression shows how PIN 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 5, the coefficient for retail trading is positive; as the proportion of retail trading across firms in the sample increases, PIN increases. 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 5, are qualitatively identical to those for the overall sample.

The results of this analysis provide evidence that a firm's PIN 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 PIN. Just as described above with respect to 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 (PIN)), 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 PIN. This analysis suggests that retail trading causes changes in the probability of informed trade (PIN).

(v) Study Limitations

This study suffers from two primary 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 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

PIN, on the one hand, and retail trading, on the other hand, 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.[177] The distinction is meaningful to the extent we hope to draw conclusions about the investment capacity of retail investors to inform securities regulation policy.

V. Discussion and Conclusion

Retail investors are abandoning the equity capital markets. This has significant implications not only for capital formation and investor protection, but also for market efficiency. This Article demonstrates that higher levels of trading by retail investors are associated with more firm-specific return variation (lower R^2) and higher probability of informed trading (PIN)[178] and also provides some evidence that the relationship between retail trading, on the one hand, and R^2 or PIN, on the other hand, is causal. The findings of this study, therefore, suggest that, contrary to the received wisdom, the presence of retail investors, as a group, in equity markets *increases* share price accuracy and market efficiency.

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 can be valuable in helping the market set prices.[179] 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. This is an

---

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

[178] For 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.

[179] *See, e.g.*, JAMES SUROWIECKI, THE WISDOM OF CROWDS: WHY THE MANY ARE SMARTER THAN THE FEW AND HOW COLLECTIVE WISDOM SHAPES BUSINESS, ECONOMIES, SOCIETIES AND NATIONS (2004); La Blanc & Rachlinski, *supra* note 65, at 542; Henry Manne, *Remarks on the Lewis & Clark Law School Business Law Forum: Behavioral Analysis of Corporate Law: Instruction or Distraction?*, 10 LEWIS & CLARK L. REV. 169 (2006);.

example of "wisdom of the crowd."[180] The findings of Jackson[181] and Choe, Kho and Stulz,[182] described in Part II, are consistent with this view.

Moreover, as I have argued elsewhere,[183] institutional trades motivated by liquidity are eclipsing trades driven by information affecting fundamental firm value. As prominent hedge fund manager and former Wall Street risk manager Richard Bookstaber notes, information is no longer the primary driver of trading activity on today's debt and equity capital markets.[184] Instead, it is the need for liquidity.[185] Bookstaber provides a number of examples of types of trades that are liquidity-induced and not driven by fundamental information.[186] Researchers studying Hong Kong-listed securities trading on the London Stock Exchange find that trading is driven largely by liquidity, not information.[187] It, therefore, is not inconceivable that prices in U.S. equity markets can be disproportionately moved by factors unrelated to fundamental value. Much of institutional trading, which affects market prices, can be characterized fairly as "noise," as it does not serve to impound relevant information into stock prices. Thus, retail traders, who do not have the same ever-present liquidity needs of institutions, are more likely to trade based on information and enhance allocative efficiency.

The instant study, therefore, not only calls into question the need for policy changes to restrict the access of retail traders, but also provides support for efforts to protect individual investors and increase their market participation. It is well-known that retail trading can add market liquidity, particularly in small firm stocks, for which liquidity often is a significant problem. Moreover, there is evidence that retail investors perform unique market functions, as they provide liquidity to institutional investors who require immediacy in trade execution.[188] 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.

---

[180] *See generally* SUROWIECKI, *supra* note 179.

[181] Jackson, *supra* note 83.

[182] Choe et al., *supra* note 87.

[183] Alicia Davis Evans, *A Requiem for the Retail Investor?*, 95 VA. L. REV. 1105, 1121 (2009).

[184] RICHARD BOOKSTABER, A DEMON OF OUR OWN DESIGN: MARKETS, HEDGE FUNDS, AND THE PERILS OF FINANCIAL INNOVATION 182 (2007).

[185] *Id.*

[186] *Id.* at 182-84.

[187] Sumit Agarwal et al., *Where Does Price Discovery Occur for Stocks Traded in Multiple Markets? Evidence from Hong Kong and London*, 26 J. INT'L MONEY & FIN. 46, 62 (2007).

[188] *See* Ron Kaniel et al., *Individual Investor Trading and Stock Returns*, 63 J. OF FIN. 273 (2008).

Appendix A

Interpreting R^2

In this study, I use low R^2 as a metric of share price accuracy. However, that interpretation is controversial. Hou, Peng and Xiong,[189] for example, question the use of low R^2 as a measure of informational efficiency.[190] They base this view on the results of independent empirical analysis and also point to other studies, such as those described below, as evidence consistent with their interpretation of R^2.

Chan and Hameed 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).[191] Based on these results, Chan and Hameed conclude that the conventional wisdom that research analysts produce firm-specific information is incorrect and that what analysts actually produce is market-wide information.[192] Hou, Peng and Xiong 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.[193]

The work of Veldkamp[194] attempts to bridge the gap between these two opposing conclusions. Veldkamp sets forth a model in which "investors purchase information that generates comovement."[195] Consistent with this model, she argues, is the finding by Hameed, Morck and Yeung[196] that firms (after controlling for size) with more analyst coverage tend to have fundamentals that predict other firms' fun-

---

[189] Kewei Hou et al., R^2 and Price Inefficiency (2006) (unpublished manuscript) (on file with author).

[190] Another potential criticism of R^2 is that it 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, as described previously, the level of R^2 reflecting informational efficiency is an interpretation that has been used extensively in recent finance scholarship.

[191] Kalok Chan & Allaudeen Hameed, *Stock Price Synchronicity and Analyst Coverage in Emerging Markets*, 80 J. FIN. ECON. 115 (2006). This result is consistent with that of a similar study of firms in the United States. *See* Joseph D. Piotroski & Darren T. Roulstone, *The Influence of Analysts, Institutional Investors, and Insiders on the Incorporation of Market, Industry and Firm-Specific Information into Stock Prices*, 79 ACCT. REV. 1119 (2004).

[192] Chan & Hameed, *supra* note 191.

[193] Hou et al., *supra* note 189.

[194] Laura L. Veldkamp, *Information Markets and the Comovement of Asset Prices*, 73 REV. ECON. STUD. 823 (2006).

[195] *Id.* at 823.

[196] Allaudeen Hameed et al., Information Markets, Analysts, and Comovement in Stock Returns (2005) (unpublished manuscript) (on file with author).

damentals. 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 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[197] cite a study performed by Ashbaugh-Skaife, Gassen and LaFond,[198] 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 find that firms with low R^2's are more likely to have accounting-based return anomalies, poor earnings quality, and weak fundamentals. [199] 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 find evidence that stocks with lower R^2's exhibit what the researchers term "overreaction-driven price momentum" and more long run price-reversals.[200]

Finally, legal scholar Ferrell reports the change of R^2 in over-the-counter stocks (OTC) after passage of the 1964 Securities Acts Amendments in the United States.[201] 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 disclo-

[197] Hou et al., *supra* note 189.

[198] Ashbaugh-Skaife, *supra* note 137.

[199] Siew Hong Teoh et al., R-square: Noise or Firm-Specific Information? (2006) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=926948.

[200] Hou et al., *supra* note 189. 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. *Id.*

[201] Allen Ferrell, *Mandated Disclosure and Stock Returns: Evidence from the Over-the-Counter Market* (Harv. L. & Econ. Discussion Paper No. 453, Dec. 2003), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=500123.

sure from OTC market firms before the amendments. Ferrell reports that, in 1963, the Securities and Exchange Commission (SEC) completed a report on the state of securities regulation. In its report, the SEC found that, of a random sample composing 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.

Lee and Liu attempt to reconcile the two competing interpretations of R^2 (i.e., low R^2 is a sign of informational efficiency or, alternatively, low R^2 is a sign of informational inefficiency).[202] 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,[203] 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.[204] 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.[205]

Dasgupta, Gan and Gao also set forth a theory that is highly persuasive in reconciling the competing views on R^2.[206] 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

---

[202] Dong Wook Lee & Mark H. Liu, Does More Information in Stock Price Lead to Greater or Smaller Idiosyncratic Return Volatility? (2007) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=887026.

[203] 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 the PIN market microstructure model. Lee & Liu, *supra* note 202. Lee and Liu acknowledge that interpreting lower values of this metric as a sign of greater price informativeness is controversial. *Id.* Indeed, it is not the dominant interpretation or the one I adopt in this study.

[204] The researchers rely on the information environment rather than the metric of informativeness used by Durnev et al., *supra* note 39 (that is, how well the price predicts future earnings) because of data availability and tractability given their research design.

[205] See Teoh et al., *supra* note 199, for evidence that calls into question certain elements of Lee and Liu's theory of the U-shaped relationship between price informativeness and idiosyncratic volatility.

[206] Dasgupta et al., *supra* note 161.

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 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").[207] 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.[208]

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 this important debate. Given the importance to securities regulators of having a way of measuring the effects of their policies on allocative efficiency, finding a reliable measure of share price accuracy is important. First, as described above, Dasgupta, Gan and Gao argue that the quality of information environment is important in interpreting R^2.[209] 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, 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.[210]

---

[207] This analysis assumes, of course, that the market would be able to incorporate the information into the price accurately.

[208] Dasgupta, Gan and Gao 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. *Id.*

[209] *Id.*

[210] Dasgupta, Gan, and Gao 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). *Id.* One may think of "time-invariant" characteristics as related to the "base" of information or general

The results[211] 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^2s. All of these factors, as discussed previously, are likely to contribute to a high-quality information environment and make firms operating in this environment less likely to experience non-market-related stock movements during the study.[212] 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^2s. 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. Firms with characteristics that are consistent with a high-quality information environment tend to have higher R^2s. Conversely, firms with traits associated with low-quality information environments tend to have lower R^2s. These factors all relate to the "base" of information about a firm.

As Durnev et al. 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.[213] 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

---

information environment and "time-variant" characteristics as related to the "flow" of information.

[211] Note that this Part 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, *supra* note 145, 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.

[212] Recall the argument of Dasgupta, Gan, and Gao. 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. Dasgupta et al., *supra* note 161.

[213] Durnev et al., *supra* note 39.

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 firm-specific news is associated with a firm's stock returns tracking the broader market less closely is unsurprising.[214] News releases are a clear example of an item affecting the "flow" of information into stock prices.[215]

As noted above, retail trading levels also affect the "flow" of information in stock prices. Indeed, there is evidence that suggests that traders are likely to be *more* influential on the incorporation of fundamentals into stock prices than is the release of news items. Roll finds that individual firm-specific stock price movements generally are not correlated with an identifiable public dissemination of news.[216] 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. A higher proportion of retail trading generally correlates with a greater number of individuals influencing asset prices.[217] 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, 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 this is a reasonable interpretation, as is the implication that trading by retail investors increases share price accuracy.

Finally, this study also studies the relationship between PIN and retail trad-

[214] Of course, if Dasgupta, Gan, and Gao 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.

[215] 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.

[216] Roll, *supra* note 92.

[217] 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.

ing. High PIN is a widely used measure of stock price informativeness in the finance literature. In this study, I demonstrate that both high PIN and low R^2 are positively associated with the level of retail trading. This result is highly suggestive of low R^2 representing greater share price accuracy, at least in certain contexts.

Table 1
Summary Statistics

Dependent Variable		Overall Sample	Top Half	Bottom Half
R2	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
PIN	Mean	0.128	0.102	0.153
	Std. Dev.	0.051	0.039	0.048
	Minimum	0.000	0.000	0.051
	Maximum	0.345	0.236	0.345
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
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**	1														
	(0.000)															
Size, vol. & res.	-0.442**	0.094**	1													
	(0.000)	(0.002)														
Insider ownrshp	0.019	0.130**	-0.248**	1												
	(0.527)	(0.000)	(0.000)													
5% owner ownrshp.	0.073*	0.104**	0.309**	0.230**	1											
	(0.015)	(0.001)	(0.000)	(0.000)												
Diverse	-0.040	-0.058+	0.078**	-0.064*	-0.106**	1										
	(0.178)	(0.053)	(0.008)	(0.032)	(0.000)											
News days	-0.148**	-0.015	0.595**	-0.204**	-0.208**	0.145**	1									
	(0.000)	(0.511)	(0.000)	(0.000)	(0.000)	(0.000)										
Age	-0.115**	-0.005	0.129**	-0.082**	-0.232**	0.284**	0.131**	1								
	(0.000)	(0.876)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)									
Volatility	0.360**	0.029	-0.247**	0.092**	0.259**	-0.173**	-0.107**	-0.228**	1							
	(0.000)	(0.334)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)								
Return on equity	-0.132**	0.042	0.134**	0.019	-0.064*	-0.038	-0.022	0.024	-0.226**	1						
	(0.000)	(0.159)	(0.000)	(0.524)	(0.031)	(0.198)	(0.468)	(0.413)	(0.000)							
Dividends-to-equity	-0.099**	-0.176**	0.149**	-0.123**	-0.248**	(0.175)**	0.132**	0.368**	-0.383**	0.148**	1					
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)						
Tang assets to total assets	0.094**	-0.090**	0.040	-0.137**	-0.075*	0.011	-0.019	0.146**	0.104**	-0.057+	0.097**	1				
	(0.002)	(0.003)	(0.175)	(0.000)	(0.011)	(0.716)	(0.531)	(0.000)	(0.000)	(0.057)	(0.001)					
R&D to assets	-0.047	(0.037)	0.049+	-0.087**	-0.046	0.043	0.097**	0.118**	0.030	-0.068*	(0.006)	-0.125**	1			
	(0.111)	(0.212)	(0.098)	(0.003)	(0.124)	(0.153)	(0.001)	(0.000)	(0.322)	(0.023)	(0.833)	(0.000)				
Book-to-mkt. equity	0.141**	-0.049	-0.275**	-0.023	0.094**	-0.004	-0.095**	-0.019	-0.001	-0.232**	0.026	(0.003)	-0.151**	1		
	(0.000)	(0.101)	(0.000)	(0.438)	(0.002)	(0.888)	(0.002)	(0.521)	(0.987)	(0.000)	(0.390)	(0.918)	(0.000)			
External finance	0.095**	0.033	0.023	0.043	-0.017	-0.023	-0.015	-0.103**	0.097**	-0.020	-0.062*	-0.071	-0.051+	-0.081**	1	
	(0.001)	(0.275)	(0.449)	(0.152)	(0.580)	(0.445)	(0.626)	(0.001)	(0.001)	(0.497)	(0.037)	(0.017)*	(0.086)	(0.006)		
Sales growth	0.116**	0.016	0.080**	0.041	-0.038	-0.067*	-0.050+	-0.127**	0.161**	0.152**	-0.101**	0.030	-0.016	-0.148**	0.482**	1
	(0.000)	(0.381)	(0.007)	(0.167)	(0.203)	(0.025)	(0.092)	(0.000)	(0.000)	(0.000)	(0.001)	(0.320)	(0.394)	(0.000)	(0.000)	

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

	R^2			PIN		
	Overall Sample	**Top Half**	**Bottom Half**	**Overall Sample**	**Top Half**	**Bottom Half**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	-1.681** (-5.65) [.297]	-2.399** (-5.90) [.407]	-.916* (-2.01) [.457]	.145** (12.84) [.011]	.166** (6.88) [.024]	.133** (7.56) [.018]
Retail trading	-.112** (-3.07) [.036]	-.028 (-0.56) [.050]	-.173** (-3.39) [.051]	.011** (8.57) [.001]	.013** (6.71) [.002]	.010** (5.99) [.002]
Institutional ownership	.202 (1.07) [.189]	-.039 (-0.15) [.258]	.175 (0.67) [.262]	-.010 (-1.48) [.006]	.003 (0.29) [.012]	-.018* (-2.28) [.008]
Size, volume & research	.171** (6.99) [.024]	.219** (5.77) [.038]	.197** (3.93) [.050]	-.012** (-13.23) [.001]	-.012** (-9.70) [.001]	-.012** (-8.78) [.001]
Insider ownership	.014 (0.95) [.014]	-.002 (-0.11) [.018]	.033 (1.50) [.022]	-.002** (-3.12) [.001]	-.001 (-0.78) [.001]	-.002** (-3.31) [.001]
5% owner ownership	-.423** (-3.19) [.133]	-.476* (-2.46) [.193]	-.416* (-2.32) [.179]	.017** (3.33) [.005]	.021* (2.31) [.009]	.011 (1.59) [.007]
Diverse	.031 (0.53) [.058]	.149+ (1.82) [.082]	-.074 (-0.94) [.079]	-.001 (-0.48) [.002]	-.002 (-0.61) [.003]	-.001 (-0.33) [.003]
News days	-.207** (-4.14) [.050]	-.213** (-3.28) [.065]	-.227** (-2.76) [.082]	.001 (0.78) [.002]	.001 (0.41) [.002]	.003 (1.23) [.003]
Age	.144** (4.32) [.033]	.159** (3.37) [.047]	.174** (3.41) [.051]	-.007** (-5.45) [.001]	-.011** (-2.88) [.004]	-.006* (-2.41) [.003]
Volatility	-1.268 (-1.03) [1.223]	3.855* (2.50) [1.540]	-3.945* (-2.43) [1.621]	.059 (1.60) [.037]	.098+ (1.80) [.055]	-.013 (-0.24) [.054]
Return on equity	-.189 (-1.33) [.142]	-.162 (-0.84) [.193]	-.146 (-0.71) [.207]	.002 (0.28) [.007]	.005 (0.45) [.010]	.001 (0.06) [.010]

Dividends-to-equity	.039** (3.16) [.012]	.040* (2.42) [.016]	.036* (2.09) [.017]	-.002** (-4.79) [.000]	-.003** (-3.63) [.001]	-.002** (-3.17) [.001]
Tangible assets to total assets	.638** (7.38) [.086]	.615** (5.60) [.110]	.498** (3.96) [.126]	-.003 (-0.93) [.003]	-.007+ (-1.77) [.004]	.004 (0.92) [.004]
R&D to assets	-.062** (-6.09) [.010]	-.080** (-5.92) [.014]	-.038* (-2.59) [.015]	.001* (2.16) [.000]	.001* (2.31) [.001]	.000 (0.17) [.001]
Book-to-market equity	-.210* (-2.03) [.103]	.124 (0.89) [.140]	-.381** (-2.96) [.129]	.011** (2.95) [.004]	.004 (0.77) [.005]	.017** (3.34) [.005]
External finance	.207 (0.90) [.229]	.486 (1.60) [.304]	-.046 (-0.15) [.310]	.002 (0.28) [.008]	-.005 (-0.38) [.015]	.008 (0.73) [.011]
Sales growth	.594** (3.18) [.187]	.530* (2.25) [.235]	.505+ (1.83) [.276]	-.026** (-4.02) [.006]	-.027* (-2.48) [.011]	-.025** (-3.00) [.008]
F-statistic	25.33	18.54	10.43	64.70	38.90	24.11
R^2	0.32	0.33	0.32	0.56	0.62	0.49
No. of observations	1129	565	564	1126	563	563

NOTE. — The table above reports the results of ordinary least squares regression of R^2 and retail trading and of PIN on retail trading 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 (-0.26) [.033]	.132** (2.86) [.046]	-.043 (-0.98) [.044]	.092 (1.03) [.090]	-.032 (-0.66) [.048]	.133* (2.14) [.062]
Volatility	6.77** (7.49) [.905]	7.729** (3.06) [2.526]	6.712** (4.88) [1.376]	16.778** (3.00) [5.601]	6.638** (5.85) [1.134]	3.859 (1.26) [3.074]
Return on equity	.134 (1.09) [.124]	-.169 (-0.95) [.178]	-.211 (-1.43) [.148]	-.626+ (-1.93) [.325]	.205 (1.29) [.159]	-.086 (-0.38) [.226]

Dividends-to-equity	.004 (0.35) [.011]	.023 (1.40) [.016]	.023 (1.47) [.016]	.061+ (1.83) [.033]	-.016 (-1.12) [.014]	.003 (0.14) [.022]
Tangible assets to total assets	.107 (1.55) [.069]	.781** (6.87) [.114]	.108 (1.15) [.094]	.855** (3.97) [.215]	.069 (0.70) [.099]	.573** (3.88) [.148]
R&D to assets	-.015* (-1.52) [.010]	-.071** (-4.94) [.014]	-.036** (-2.80) [.013]	-.138** (-4.32) [.032]	-.006 (-0.41) [.015]	-.036+ (-1.92) [.019]
Book-to-market equity	-.030 (-0.36) [.081]	-.088 (-0.75) [.117]	-.406** (-3.06) [.133]	-.472 (-1.48) [.319]	.053 (0.55) [.095]	-.181 (-1.27) [.142]
External finance	.398* (2.07) [.193]	.640* (1.97) [.325]	.739* (2.57) [.288]	1.852* (2.40) [.772]	.279 (1.22) [.228]	.196 (0.51) [.381]
Sales growth	.799** (4.77) [.168]	1.200** (4.27) [.281]	.953** (3.70) [.258]	1.967** (2.80) [.702]	.774** (3.72) [.208]	.988** (2.81) [.351]
Stock price	-.342** (-6.91) [.049]		-.226** (-3.15) [.072]		-.378** (-5.53) [.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.

Table 5

Instrumental Variable Estimates of the Relationship between PIN and Retail Trading Activity

	Overall Sample		*Top Half*		*Bottom Half*	
	First Stage Retail Trading	*Second Stage* PIN	*First Stage* Retail Trading	*Second Stage* PIN	*First Stage* Retail Trading	*Second Stage* PIN
	1.257**	.149**	1.859**	.138**	1.310**	.120**
Constant	(3.58)	(8.22)	(2.83)	(3.98)	(2.72)	(4.06)
	[.351]	[.018]	[.658]	[.035]	[.514]	[.030]
		.057**		.057**		.063**
Retail trading		(6.36)		(4.97)		(3.98)
		[.009]		[.011]		[.016]
	-2.230**	.101**	-2.855**	.139**	-1.818**	.085*
Institutional ownership	(-16.04)	(4.31)	(-12.84)	(3.60)	(-10.00)	(2.58)
	[.139]	[.023]	[.222]	[.039]	[.182]	[.033]
	-.257**	.001	-.248**	-.000	-.269**	.004
Size, volume & research	(-12.49)	(0.44)	(-8.87)	(-0.03)	(-8.71)	(0.83)
	[.021]	[.003]	[.028]	[.004]	[.031]	[.005]
	-.013	-.001	-.008	.000	-.023	-.001
Insider ownership	(-1.00)	(-1.38)	(-0.35)	(0.15)	(-1.50)	(-1.09)
	[.013]	[.001]	[.022]	[.001]	[.016]	[.001]
	-.274*	.032**	-.402*	.038**	.011	.013
5% owner ownership	(-2.11)	(3.77)	(-2.01)	(2.99)	(0.07)	(1.14)
	[.130]	[.008]	[.200]	[.013]	[.173]	[.012]
	.025	-.002	.027	-.003	.035	-.002
Diverse	(0.45)	(-0.49)	(.34)	(-0.69)	(0.46)	(-0.38)
	[.055]	[.003]	[.078]	[.005]	[.077]	[.005]
	.184**	.009**	.224**	-.010*	.099	-.004
News days	(4.30)	(-2.70)	(4.41)	(-2.48)	(1.34)	(-0.70)
	[.043]	[.003]	[.051]	[.004]	[.074]	[.005]
	-.012	-.007**	-.043	-.007	-.128*	.001
Age	(-0.35)	(-3.37)	(-0.49)	(-1.41)	(-2.11)	(0.30)
	[.033]	[.002]	[.086]	[.005]	[.061]	[.005]
	6.79**	-.335**	7.531**	-.322*	6.774**	-.450**
Volatility	(7.49)	(-3.57)	(5.80)	(-2.40)	(5.37)	(-3.01)
	[.906]	[.094]	[1.298]	[.134]	[1.262]	[.149]
	.137	.001	.308	-.005	-.005	.008
Return on equity	(1.10)	(0.15)	(1.60)	(-0.61)	(-0.03)	(0.67)
	[.124]	[.007]	[.193]	[.009]	[.173]	[.011]

Dividends-to-equity	.003 (0.31) [.011]	-.001* (-2.33) [.001]	-.014 (-0.84) [.017]	-.001 (-0.91) [.001]	.017 (1.08) [.015]	-.002* (-2.26) [.001]
Tangible assets to total assets	.108 (1.56) [.069]	-.009* (-2.03) [.004]	.037 (0.40) [.092]	-.010+ (-1.83) [.006]	.163 (1.58) [.103]	-.006 (-0.83) [.008]
R&D to assets	-.014 (-1.43) [.010]	.001* (2.04) [.001]	.003 (0.24) [.013]	.001 (1.21) [.001]	-.034* (-2.22) [.015]	.002 (1.47) [.001]
Book-to-market equity	-.029 (-0.35) [.082]	.005 (1.07) [.005]	.101 (0.88) [.115]	-.006 (-0.85) [.007]	-.077 (-0.69) [.111]	.013+ (1.90) [.007]
External finance	.397* (2.06) [.193]	-.017 (-1.43) [.012]	.576* (2.07) [.278]	-.034+ (-1.87) [.018]	.263 (1.05) [.250]	-.006 (-0.40) [.015]
Sales growth	.799** (4.76) [.168]	-.052** (-4.74) [.011]	.740** (3.04) [.244]	-.045** (-2.78) [.016]	.928** (4.18) [.222]	-.066** (-3.54) [.019]
Stock price	-.340** (-6.88) [.049]		-.360** (-5.29) [.068]		-.307** (-4.29) [.071]	
F-statistic	65.72		42.47		32.74	
No. of observations:	1126		563		563	

NOTE. – The table above reports the results of a regression of PIN 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.