Centre for
Economic
and Financial
Research
at
New Economic
School

September 2014

Information Efficiency and Firm-Specific Return Variation

Patrick J. Kelly

Working Paper No 208

CEFIR / NES Working Paper series

Information Efficiency and Firm-Specific Return Variation*

Patrick J. Kelly
New Economic School

Version: September 2014

* Special thanks to John Griffin, Spencer Martin, Jeff Coles, and Federico Nardari. I would also like to thank Chris Anderson, Ana Balcarcel, George Cashman, Art Durnev, Paul Irvine, Jennifer Juergens, Lidia Kelly, Paul Koch, Olga Kuzmina, Laura Lindsey, Dmitry Makarov, Felix Meschke, Laura Tuttle, Sriram Villupuram and seminar participants at Arizona State University, the Bank of Canada, Drexel University, the University of Kansas, the Securities and Exchange Commission, the University of South Florida, and Tulane University, for their insightful comments. I/B/E/S data are generously provided by Carr Bettis and Camelback Research.
Patrick J. Kelly is an Associate Professor in the Department of Finance at the New Economic School, 100 Novaya Street, Skolkovo, 143025 Moscow, Russia. E-mail: pkelly@nes.ru tel: +7 (495) 956.95.08

Information Efficiency and Firm-Specific Return Variation

Abstract

Reasoning that private firm-specific information causes firm-specific return variation that drives down market-model R^2s, Morck, Yeung, and Yu (2000) begin a large body of research which interprets R^2 as an inverse measure of price informativeness. Low R^2s or "synchronicity," as it is called in this literature, signal that prices more efficiently incorporate private firm-specific information, and high R^2s indicate less. For this to be true, we would expect that low-R^2 stocks have characteristics that facilitate private informed trade, i.e. lower information costs and fewer impediments to arbitrage. However, in this paper we document the opposite: Low-R^2 stocks are small, young, and followed by few analysts, and have high bid-ask spreads, high price impact, greater short-sale constraints and are infrequently traded. In fact, microstructure measures suggest that private-information events are less likely for low-R^2 stocks than high, and that differences in R^2 are driven as much by firm-specific volatility on days without private news as by firm-specific volatility on days with private news. These results call into question prior research using R^2 to measure the information content of stock prices.

1. Introduction

The greatest portion of return variation is unexplained by current asset pricing models. On average, standard models account for only 17% of daily return variation and 29% of monthly. Observing a similar lack of model fit, after controlling for exposure to systematic risk, industry-specific factors, and the occurrence of value-relevant public information, Roll (1988) concludes that the majority of returns are explained either by private information or a "frenzy" unrelated to specific information.

Work by Morck, Yeung and Yu (2000, 2013), Durnev, Morck, Yeung and Zarowin (2003) and Durnev, Morck, and Yeung (2004) provides evidence for the private information explanation. This research shows that a low market-model R^2 (called "synchronicity" in this literature) is associated with fewer legal and regulatory impediments to informed trade across countries, and within U.S. markets more efficient corporate investment and returns which are more sensitive to future earnings growth. This evidence is consistent with the notion that low R^2s result from the incorporation of private, firm-specific information which makes prices more informationally efficient and leads these papers to posit that market-model R^2 is an inverse measure of "price informativeness" or information efficiency.[1]

This paper examines this contention that R^2 is an inverse measure of information efficiency. We first look at the information environment surrounding stocks as categorized by this widely used measure of information efficiency[2] and find that the information environment is particularly poor for low-R^2 stocks. This alone would seem to contradict the proposition that R^2 varies inversely with information efficiency. Second, we use microstructure measures to examine the impact private information-based trade on idiosyncratic volatility and poor model fit (R^2). In doing so, we provide evidence that low R^2s are associated with both private information and sources unrelated to specific

[1]A number of studies provide corroborating evidence. When average R^2 is low, capital markets are more open (Li, Morck, Yang, and Yeung, 2004), short sales are less constrained (Bris, Goetzmann, and Zhu, 2007), capital is better allocated, and government ownership in the economy is less (Wurgler, 2000).

[2] Cited by 348 published papers according to Thomson Reuters' Web of Science on September 9, 2014.

information. Specifically, we estimate the probably of private information arrival on each day and show that the average day with private information does in fact have higher idiosyncratic volatility than the average day without, especially for low-R^2 (low synchronicity) stocks, consistent with the arguments in Morck, Yeung and Yu (2000) and other papers. However, days with private information are infrequent, occurring only 30% of the time for high-R^2 stocks, and only 15% of the time for low R^2. As a result, when aggregated over the course of a year, most idiosyncratic return occurs on days without private information simply because there are more days without private news. The fact that this is particularly true for low-R^2 stocks suggests that R^2 is a poor measure of private information incorporation.

The idea that volatility might reflect information incorporation is not new. French and Roll (1986), note that the key distinction between public and private information is that public information affects prices the moment it becomes known, while private information is only revealed through trading. French and Roll (1986) along with Barclay, Litzenberger, and Warner (1990), and Jones, Kaul, and Lipson (1994) find differences in volatility during trading and non-trading hours. Their evidence suggests that the greater portion of return volatility is due to the activity of private-information driven traders.

On the other hand, Shiller (1981), LeRoy and Porter (1981), and West (1988) present evidence that stock returns are significantly more volatile than the random arrival of new value relevant information would permit in an efficient market. They show that rapid information incorporation results in lower volatility, not higher, because changes in expected firm value that are incorporated in a stock's price sooner are more heavily discounted. West (1988) generalizes these models to show that return variance is greater anytime the information set on which expectations are based is a subset of all available information.

From a microstructure perspective informationally efficient prices result from traders who arbitrage their information advantage (Grossman and Stiglitz, 1980, Glosten and Milgrom, 1985, and Kyle, 1985). In these models low information cost and high liquidity promote the acquisition of information and its incorporation into stock prices.

In this paper we first examine the association between model fit (R^2) and measures of information cost and liquidity. We find that low-R^2 stocks are those with higher information costs and less liquidity, inconsistent with the notion that poor model fit (low R^2 or high idiosyncratic volatility) is predominantly the result of private informed trade; rather, low-R^2 stocks are those with the least trade and with the greatest impediments to informed trade. We show that a market-model R^2 (idiosyncratic volatility) is positively (negatively) associated with the quality of the information environment, consistent with the evidence of Shiller (1981), LeRoy and Porter (1981), and West (1988).[3]

In addition, we look at improvements in the quality of the information environment through time, using initiation of analyst coverage as a proxy for improvement for the quality of the information environment. Changes are most consistent with mean reversion, as low-R^2 stocks tend to have higher R^2s following initiation of analyst coverage and high-R^2 stocks tend to be lower following initiation. Over the entire sample there is an economically small increase in R^2 associated with the initiation of analyst coverage. Nonetheless this small increase in R^2 as the information environment improves is inconsistent with the notion that R^2 is an inverse measure of efficiency.

In this study we also utilize daily measures of private information arrival, derived from the microstructure model of Easley, Kiefer and O'Hara (1997) (hereafter, EKO97), to estimate the probability of private information events, the degree of information asymmetry, and the level of informed and uninformed trade. Consistent with our finding that low-R^2 stocks face greater

---

[3] Umlauf (1993), Jones and Seguin (1997), and Bessembinder and Rath (2008) also find that reductions in trading costs, which they argue facilitate informed trade, are associated reductions in stock volatility.

impediments to informed trade, we find that low-R^2 stocks have fewer expected informed trades, fewer uniformed liquidity trades, and are subject to greater asymmetric information risk, while having a lower likelihood of a private information event. These finding suggest that a low market-model R^2 is a sign of relatively less informationally efficient pricing.

Finally, we contrast the proportion of idiosyncratic volatility ($1 - R^2$) that occurs as a result of private information. We do this in two ways, the first method attributes all idiosyncratic volatility on days with a high probability of a private information event to private information and compares it to the idiosyncratic volatility on days with a low probability. As noted earlier, we find that while private information events are quite rare for the lowest R^2 stocks, private information does significantly impact market-model R^2s. Roughly half a stock's idiosyncratic volatility occurs on days with a high probability of a private information event. The second method uses the probability of a private information event as a regressor and examines the improvement in model fit. We find its impact is greatest for the lowest R^2 stocks, however, it explains at most an additional 13% of stocks' return variation above and beyond that explained by other sources of return comovement: size, book-to-market, momentum and liquidity factors (including controls for infrequent trading and bid-ask bounce).

Together these findings suggest a response to the question implied by Roll (1988): is poor model fit a function of private information incorporation or to a "frenzy" unrelated to information? The answer is both. Privately informed trade does cause prices to deviate from the market-model expected return, however there remain other sources unrelated to firm-specific information as well. Importantly, the proportion of return not associated with specific information is sufficiently large to warrant caution when using market-model R^2 or synchronicity as an inverse measure of information efficiency.

Because the focus on this paper is on whether market-model R^2 might be an inverse measure of private information incorporation and information efficiency, our results do not preclude many earlier findings regarding possible sources of idiosyncratic return variation. Earlier work has shown that idiosyncratic volatility has increased over time in the US (Campbell, Lettau, Malkiel and Xu, 2001) and around the world (Jin and Myers, 2006, and Li, Morck, Yang and Yeung, 2004). Others have argued that idiosyncratic volatility is a function of volatile fundamentals (Wei and Zhang, 2006, and Bartram, Brown and Stulz, 2012), whether it is due to age (Pastor and Veronesi, 2003, Fink, Fink, Grullon and Weston, 2010), greater firm focus and leverage (Dennis and Strickland, 2009), increased competition (Irvine and Pontiff, 2003) and growth options (Cao, Simin and Zhao, 2008). Lee and Liu (2011) and Xing and Anderson (2011) show R^2 can be either high or low in good quality information environments, depending on the firm characteristics for which a test controls. Furthermore, Bartram, Brown and Stulz (2012) points to general problems using R^2, because differences in systematic risk can drive differences in R^2 as much as differences in idiosyncratic risk can.

Our findings are consistent with and complement several recent papers.[4] Gassen, LaFond, Skaife and Veenman (2014) provides compelling evidence that much of the differences in R^2 both within and across countries is driven my differences in liquidity, subverting findings in prior research suggesting that stocks' low R^2s result from transparent information environments. Hou, Xiong, and Peng (2006) provide evidence consistent with greater pronounced overreaction-driven price momentum among low-R^2 stocks. Dasgupta, Gan and Gao (2010) provide both theory and evidence that R^2 can increase when transparency improves.

[4] In research subsequent to this one, Li, Rajgopal, and Venkatachalam (2013, forthcoming) confirm the findings of this paper using many of the same or similar measures.

The remainder of the paper proceeds as follows: Section 2 reviews the Data and Methodology. Section 3 investigates impediments to informed trade as they relate to R^2. Section 4 examines the relation between impediments to informed trade and market-model fit. Section 5 concludes.

2. Data and Methodology

In this paper we use two methods to assess the arrival of information in stock prices. First, we examine the association between market-model R^2 and impediments to informed trade for all NYSE, AMEX and NASDAQ listed stocks. Second, we estimate the probability of private information arrival based on a microstructure model by Easley, Kiefer, and O'Hara (1997). We use the derived measures to directly estimate the impact of private information on returns. For practical and theoretical reasons described below, we limit this analysis to NYSE-listed stocks from 1993 through 2002. In this section we describe the data and methodology used to derive R^2 and the measures of private information and impediments to informed trade.

2.1. Return Based Measures and Impediments to Informed Trade

In this section, we describe the measures of the information environment characteristics used to proxy for impediments to informed trade as well as the data and procedures to calculate them. Data are from the Center for Research in Security Prices (CRSP), Thomson Financial and I/B/E/S. In order to be included in this study, we require a security to have data available to calculate R^2, size, age, lagged analyst count, lagged change in breadth, estimated trading costs, illiquidity, and volume. This requirement restricts the dataset to the period from 1983. Our data stop in 2002. In most analyses market-model R^2 is calculated following Durnev, Morck, Yeung and Zarowin (2003) and Durnev, Morck and Yeung (2004).[5] Each year using all common ordinary shares listed on NYSE, AMEX or NASDAQ with 52 weeks of returns, we regress weekly Wednesday-to-Wednesday

[5] Unlike Durnev et al. (2003, 2004) we do not exclude utilities and financial companies. This is done for comparability across parts of this study. Results are qualitatively the same with or without utilities and financial companies.

individual stock returns on value-weighted market return and value-weighted two-digit SIC code industry returns. We include the value-weighted industry returns in order to account for non-systematic returns due to industry related factors. The regression is as follows:

$$R_{i,t} = \alpha_i + \beta_{Mkt,i,t} R_{Mkt,t} + \beta_{Ind_i i,t} R_{Ind_i \neq i,t} + \varepsilon_{i,t}, \quad (1)$$

where $R_{i,t}$ is the total return on individual stock *i*, $R_{Mkt,t}$ is the value-weighted market return, and $R_{Indi \neq i,t}$ is the value-weighted two-digit SIC industry return excluding firm *i*. The R^2s used for the purpose of creating R^2 portfolios are from the weekly regressions. In other analyses examining the impact of private information on returns in section 4, to allow the comparison of daily, weekly and monthly returns, we use 5-year non-overlapping periods. Details are described in the discussion of data below.

2.1.1. Size, Age, Turnover, Volume, and Illiquidity

Size and age are calculated at the end of each December. Size is price times shares outstanding. Age is measured as the number of years since the firm first appeared on the CRSP monthly tapes.[6] Volume and illiquidity are measures contemporaneous with the market-model R^2. The percentage of zero-volume days is calculated as the percentage of days with non-missing price data where the volume is reported as zero by CRSP. Illiquidity is measured following Amihud (2002) and is the average over the year of the absolute daily return divided by the daily dollar volume of trade.

2.1.2. Lesmond, Ogden, and Trzcinka (1999) Trading Costs

Lesmond, Ogden and Trzcinka (1999) propose a model of trading costs which recognizes that the fundamental value of an asset is continuous while, due to trading frictions, the realization is

[6]CRSP began covering NASDAQ stocks in 1973. Since the majority of low-R^2 stocks are listed on NASDAQ, the correlation between Age and R^2 is arguably biased upward. Examining the correlation between Age and R^2 for NYSE alone results in a correlation coefficient of .17, versus the 0.29 correlation in Table 2.

not.[7] Measured returns of zero imply that the transaction costs are higher than any change in the fundamental value of the underlying asset. Observing the magnitude of returns needed to obtain a measurable non-zero return is indicative of the trading costs. Measured returns are the difference between true returns and the threshold trading cost. We follow Lesmond, Ogden and Trzcinka (1999) in the estimation and calculation of trading costs as the difference between the upper and lower thresholds their model estimates.

2.1.3. Breadth of Ownership

Data on the holdings of large institutions are from the 13f filings with the U.S. Securities and Exchange Commission (SEC) and distributed by Thomson (CDA/Spectrum). Each December institutional ownership is calculated as the percentage of shares outstanding that are held by institutions with an asset value greater than a floating threshold.[8] Breadth of ownership is calculated as the number of institutions with assets greater than the floating threshold holding a stock. Change in breadth of ownership is calculated following Chen, Hong, and Stein (2002). When calculating the change in breadth of ownership for a stock at time t the count of managers is limited to those who are holding any stock in time t-1. The change is the difference between the number of institutions holding a stock in time t and those holding in time t-1 divided by number of managers in time t-1. Data on institutional ownership are available from 1980.

2.1.4. Analyst Coverage

Data on analyst forecasts are from I/B/E/S.[9] Analyst coverage is the number of unique analysts issuing earnings forecasts during a given year for a given stock, where the entry date of the

[7] The model is inspired by Rosett (1959).

[8] Any firm with more than $100 million of securities under discretionary management must disclose holdings over $200,000 or 10,000 shares. Because the SEC does not adjust this $100 million dollar threshold, which was set in 1980, as the market has risen in value, the number of institutions required to report their holdings grows as the market rises. In order to adjust for the bias toward periods with higher market returns, the $100 million threshold is adjusted following Gompers and Metrick (1998, 2001). Each quarter, the threshold is increased by the growth in an index of all shares held by institutions.

[9] We thank Carr Bettis and Camelback Research who generously provide the data.

forecast in the I/B/E/S database is considered the earnings forecast date. The percent deviation of analyst count from the annual mean is used in regressions, in order to keep the interpretation of the coefficient estimates the same across years. Analyst count data are available from 1982.

2.2. Estimating the Probability of Information Events

Easley, Kiefer, and O'Hara (1997) develop a model of the beliefs of the market maker regarding the probability that a private information event has occurred (α) and the probability that a given trade is based on private information (PIN). In the model before the start of each trading day an information event occurs with probability α. This information event is bad news with probability δ, and good news with probability (1-δ). Trades arrive according to a Poisson process through out the day, with uninformed buys and sells arriving at the rate of ε_b and ε_s, while informed trades arrive only on days with information events at the rate of μ. The market maker observes the arrival of buy and sell trades and forms an opinion about the probability of good and bad news events, as well as estimates of the level of informed and uninformed trade. Each year we estimate these beliefs using NYSE Trade and Quote (TAQ) data from 1993 through 2002. Following Easley, Hvidkjaer, and O'Hara (2002) we limit our analysis to NYSE stocks, because the market structure of NYSE most closely resembles that of the EKO97 model. Trades and quotes are matched and buy and sell trades are assigned by the algorithm suggested by Lee and Ready (1991). We follow Bessembinder (2003) when cleaning the trade and quote data.[10] In order to estimate the model parameters, we maximize the following likelihood function:

[10] Following footnote 7 of Bessembinder (2003), we eliminate trades that are in error, a correction, out of sequence, exchange acquisitions or distributions, or involve nonstandard settlement. We eliminate quotes that are non-positive, are associated with trading halts or designated order imbalances, or are non-firm.

$$
\begin{aligned}
L(\alpha, \mu, \varepsilon_b, \varepsilon_s, \delta \mid B, S) = {} & (1-\alpha)e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!} \\
& + \alpha\delta e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-(\mu+\varepsilon_s)}\frac{(\mu+\varepsilon_s)^S}{S!} \\
& + \alpha(1-\delta)e^{-(\mu+\varepsilon_b)}\frac{(\mu+\varepsilon_b)^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!},
\end{aligned}
\tag{2}
$$

where *B* and *S* are the number of trades signed buy and sell respectively and the remaining are the estimates of the arrival rate of informed and uninformed trading described above.[11]

3. Impediments to Informed Trade and Market-model R^2

3.1. Choice of Measures

In this section, we discuss the motivation for the variables chosen to characterize the impediments to informed trade. The microstructure literature of Grossman (1976), Grossman and Stiglitz (1980) and Kyle (1985) suggest three interrelated costs play a role: information costs, explicit trading costs, and liquidity.

In a frictionless world with rational agents, if all value-relevant information were public, then prices would be informationally efficient. With costly private information, perfect informational efficiency is unattainable (Grossman and Stiglitz, 1980); however, by admitting non-information driven (noise) traders, some information can be imparted to the market. The informed trade can "hide" among liquidity (noise) trades so that prices do not adjust immediately upon an informed agent's decision to trade, which could eliminate potential profit (Grossman, 1976, Grossman and Stiglitz, 1980, and Kyle, 1985). That is to say, there are limits to arbitrage.

Even if the cost of collecting information is less than the value of information, direct trading costs (bid-ask spread, the cost of short selling, etc.) and liquidity related costs (price pressure and the

[11] We thank Soeren Hvidkjaer for advice to improve estimation procedures. See Easley, Hvdkjaer and O'Hara (2002, 2010) for additional details.

ability to conceal from the market that one is an informed trader) may impede the ability of arbitrageurs to trade on their private information, allowing mispricing to persist. It is worth noting that direct trading costs affect the profitability of arbitrage through two channels. First, wide spreads raise the arbitrageur's cost of trade, reducing the incentive to trade on private information. Second, wide spreads also raise the trading costs for the unformed. Easley, Kiefer, O'Hara and Paperman (1996) show that liquidity trading is decreasing in the size of transaction costs. Less liquidity trading, as noted above, increases the price impact of the informed trade by reducing the ability of the arbitrageur to hide trade among the trades of the uninformed.[12] In the subsections that follow, we motivate the choice of variables used to proxy for information costs, the cost of trade and liquidity.

3.1.1. Cost of Information

We use three measures to proxy for information costs: analyst coverage, size, and age. The business model for analysts takes advantage of the public-good nature of information. They pay the fixed costs of acquiring information and profit by distributing the information to investors at a price lower than the cost. In this way analysts function to lower the cost of information acquisition. One concern with the use of analyst coverage as a proxy for information costs may be that analysts have little incentive to follow firms, which are already efficiently priced and, as such, high analyst coverage may merely be indicative of inefficient pricing. For a static equilibrium this would be true, once mispricing is eliminated, no other arises – they have analyzed themselves out of business. However, in a dynamic setting, analysts have the incentive to cover firms which generate new information on a regular basis, causing frequent if temporary mispricings and the opportunity for arbitrageurs to profit from the information analysts generate. If it is private information that causes high

[12] In addition, liquidity trade must be truly random noise trading. Correlated noise trade, along the lines of DeLong, Shleifer, Summers, and Waldmann (1990), while still providing camouflage for contrarian arbitrageurs, may increase the cost and lower the profitability of arbitrage by delaying the return of market values to fundamentals (along the lines of the Dow and Gorton, 1994, "arbitrage chains" argument).

idiosyncratic volatility and poor model fit then we may find that high analyst coverage is associated with lower R^2s.[13]

Empirical evidence supports this notion that analysts precipitate the flow of information in prices. Work by Brennan, Jegadeesh, and Swaminathan (1993) finds that the returns on stocks followed by many analysts lead those of stocks followed by few analysts. Kim, Lin, and Slovin (1997), provide evidence that analysts promote the rapid incorporation of private information. According to Frankel and Li (2004) insiders profit less from their trades when there is greater analyst coverage. Hong, Lim, and Stein (2000) and Griffin and Lemmon (2002) find that momentum and book-to-market effects are concentrated in firms with few analysts and weaken dramatically with broad analyst coverage. Piotroski and Roulstone (2004) find that analysts facilitate the dissemination of industry information, more so than market or firm-specific information. Finally, analyst coverage may proxy for attention, an indirect but potentially substantial cost, that presents yet another information hurdle for the investor.

Like analyst coverage, size and age have a dual role. In the Merton (1987) sense, few investors may follow small and young firms. If traders are unaware of a stock, then they cannot discover mispricing in the stock's returns – essentially cost of information is infinite. Ho and Michaely (1988) argue that if information acquisition is more costly for small firms then, in equilibrium, investors may optimally choose to learn less about small companies. Even if the costs of learning about small stocks are no greater, the potential gains from small stock investment may be too low to justify the investment of time and money. Empirical evidence supporting the notion that size and age are associated with information costs include the following: Atiase (1985) finds that private pre-disclosure is increasing in firm size; Chemmanur and Fulghier (1999) argue that older firms are easier

[13] In unreported analyses we find that the probability of information events as estimated by the EKO97 model has a .38 Spearman rank correlation with analyst coverage, suggesting that analysts do cover firms with more private information to release.

to evaluate because there is more time to collect and process information about the firm; Grullon, Kanatas, and Weston (2004) use age to proxy for familiarity, visibility and investor recognition.

3.1.2. Costs of Trade and Liquidity

The costs of trade are frictions that impact the speed of information incorporation in prices. They can be the result of market maker overhead, compensation for liquidity provision, and adverse selection costs. There may exist a bid-ask spread, even in the absence of overhead and liquidity costs, in order to compensate the market maker for the risk of trading against an informed trader (Gloston and Milgrom, 1985, Kyle, 1985, Easley et al., 1996). When firms provide better disclosure, bid-ask spreads decrease (Helfin, Shaw, and Wild, 2005). Illiquidity and adverse selection costs are captured using several measures: the measure of trading cost developed by Lesmond, Ogden, and Trzcinka (1999); the Amihud (2002) illiquidity measure; and the percent of days with zero volume.

Short-sale constraints limit the ability to arbitrage. Diamond and Verrecchia (1987) argue that short-sale constraints reduce the speed of information incorporation in prices. As a proxy for short-sale constraints, we use change in the breadth of institutional ownership following Chen, Hong, and Stein (2002) who argue that reductions in the breadth of ownership signal that short-sale constraints are more binding and that prices are higher relative to their fundamentals.

3.2. Impediments to Informed Trade and Market-model R^2: Evidence

In the following sections, under the rational that high information costs, high trading costs and low liquidity create limits to arbitrage in which mispricing can persist, we examine the relation between model fit and impediments to trade. The central question is whether there are consistent differences in costs and liquidity based on the level of R^2. To investigate these differences, R^2 portfolio averages are presented, followed by simple correlations to understand if the patterns in the means mirror patterns at the observation level, and regressions to explore the incremental explanatory power of each of the variables and to see which impediments to trade are most closely associated with differences in R^2. The bottom line across all analyses is the same: greater information costs, greater trading costs and lower liquidity are consistently associated with low market-model R^2s and high idiosyncratic volatility. These findings are inconsistent with the notion that idiosyncratic volatility predominantly captures the incorporation of private information, and instead suggest the converse, that stocks with low market-model R^2 may be those with the greatest possibility of mispricing.

3.2.1. Sample Description

To understand the nature of the information environment surrounding stocks and how it is associated with market-model R^2, we begin with simple sorts of stocks into R^2-sorted portfolios from Eq. (1). For consistency with the prior literature (Durnev et al., 2003, 2004) these R^2s are based on annual regressions of weekly returns. As in the prior literature, we require 52 weeks of returns each year. To calculate portfolio averages we first sort all stocks into NYSE-R^2 deciles at the end of December each year (*t*). The average for each variable is calculated each year in the sample, and the time-series mean of the portfolio averages is presented in Panel A of Table 1. For size, age, analyst count and change in the breadth of institutional ownership we use data from the year before (*t*-1). For trading cost and the two illiquidity measures, we use year (*t*) averages. The number of stocks listed on each exchange is counted in December of year (*t*) and the average of these yearly counts is presented in Panel B of Table 1.

<INSERT TABLE 1 ABOUT HERE>

The dispersion of R^2 across the portfolios is large. The average R^2 for low and high portfolios is 0.025 and 0.595, respectively. Consistent with the findings of Roll (1988), firm size is monotonically increasing in R^2 (decreasing in idiosyncratic return variance). Given the small size of the low-R^2 stocks, it is not surprising to see that on average the vast majority (78%) of low-R^2 stocks trade on NASDAQ. Nearly all (84%) high-R^2 stocks trade on NYSE.

Table 1, Panel A shows that relative to high-R^2 stocks, low-R^2 stocks tend to be young, small, illiquid, and have high trading costs. There are no trades for low-R^2 stocks on 18.7% of trading days in the year. The Amihud (2002) measure of illiquidity indicates that every one hundred million dollars in trade volume results in an average 21.14% return for low-R^2 stocks, compared to a 0.02% return for high-R^2 stocks. The average roundtrip trading cost is 12.1% for low-R^2 stocks and 0.4% for high-R^2 stocks. Low-R^2 stocks receive less attention from analysts. Fewer analysts cover low-R^2 firms; on average, only one analyst covers low-R^2 stocks, whereas 23 cover high-R^2 stocks. These results are consistent with the notion that low-R^2 stocks, suffering from a poor quality information environment, face greater impediments to informed trade.

3.2.2. R^2 and Impediments to Informed Trade: Correlations

We examine simple correlations between R^2 and measures of impediments to informed trade. Table 1, Panel C presents the average of yearly cross-sectional Pearson correlation coefficients in the bottom diagonal and Spearman rank correlation coefficients in the top diagonal. In brief, the correlations are consistent with the associations between portfolio averages and the measures of impediments to trade seen in Panel A.

R^2 has high rank correlations with size (.59), analyst count(.49), trading cost (-.60), illiquidity (-.62) and the percentage of zero volume days (-.52). Except for analyst count, these correlations are noticeably weaker using the Pearson linear correlation, suggesting that there is a non-linear relation between these variables and model fit. Though lower, correlations for age (information cost) and change in breadth (relaxation of short sale constraints) are positive. Like the results from the previous section, these correlations suggest that lower information and trading costs and greater liquidity are associated with higher market-model R^2s and lower idiosyncratic volatility. However, the table also makes clear that these variables are strongly inter-related. In the next section we use regressions to examine whether each of these variables, designed to capture an aspect of impediments to informed trade, is associated with market-model R^2 in a manner consistent with the evidence in Table 1.

3.2.3. Regressions

To examine the joint relation between R^2 and impediments to informed trade, regressions are run of R^2 on information costs, trading costs and the various liquidity measures. Like the previous evidence, we find results, which suggest that low-R^2 stocks are those with the highest information and trading costs and least liquidity – inconsistent with the notion that stocks have low R^2s primarily as a result of privately informed trade.

In order to control for the fact the regressand, R^2, is bounded, we follow Durnev et al. (2004), by using the logistic transformation in lieu of R^2 itself: $ln(R^2/(1-R^2))$. This transformation is identical to the log ratio of the explained variance to unexplained variance. In addition, in order to control for the extreme leptokurtic distributions several variables exhibit, we take the natural log of the prior December market capitalization, trading costs, the Amihud (2002) illiquidity measure, and change in breadth.[14] We pool these annual data and run the regressions over the entire sample. To control for persistence in the value of company-level measures across years, we use a White (1980) heteroskedasticity correction with Rogers clustered standard errors. The White correction also controls for possible bias as a result of using the logistic transform of the regressand.[15]

To begin, in Panel A of Table 2 we report univariate regressions of the logistic-transformed R^2 on each of the measures of information costs, trading costs and liquidity. The results are consistent with the correlations found in Table 1. Lower information costs are associated with higher market-

[14]Because change in breadth can have legitimate zero values, one is added prior to taking the log. In addition, the estimate of trading costs is sometimes estimated to be zero for extremely liquid stocks (a boundary condition), in order not to exclude these from the analysis, we also add one to the trading cost measure as well. This alters the interpretation slightly, but it does not change the sign of the coefficients.

[15] When including a White (1980) heteroskedasticity correction, the resulting standard errors are unbiased – for a very interesting and useful paper on the appropriate controls for within firm and across time dependence see Petersen (2009). See Manning (1996) and Sapra (1998) for discussions of the bias induced when using a logistic transform of a regressand.

model R^2s, as are lower trading costs and greater liquidity and less tightly binding short-sale constraints. However, firm age and change in breadth (our proxy for short sale constraints) explain little of the differences in R^2; their univariate regression R^2s are only .065 and .047, respectively. In addition, the explanatory power of our trading cost estimate is completely subsumed by the percent of days with zero volume in Panel B, so much so, that in some specifications the sign on trading costs flips.

<INSERT TABLE 2 ABOUT HERE>

Notable is the relatively strong association between R^2 and market capitalization and Amihud (2002) illiquidity. Respectively, each explains 23.4% and 26.4% on its own, while the best model explains only 6% more (Model 13, 32.5%). Adding market capitalization to the regression with illiquidity, R^2 barely improves. This suggests that differences in R^2 are most closely associated with differences in liquidity (see, Gassen, LaFond, Skaife and Veenman, 2014, for a careful analysis of the impact of illiquidity on R^2).

There is a known relation between trading costs and volatility, and liquidity and volatility. To examine the possibility that the associations we find between R^2 and impediments to informed trade are driven by an association with volatility, we include Schwert (1990) volatility as a control variable in the regression in column 14. The results remain substantially unchanged.

The findings in this section show that impediments to informed trade, higher information costs, higher trading costs, and lower liquidity are associated with a lower market-model R^2. They suggest that low-R^2 stocks are more costly to arbitrage and, as such, present conditions where mispricing may persist. The results also show that Amihud (2002) illiquidity is highly correlated with differences in market-model R^2. This finding suggests an intriguing possibility: differences in market-model R^2, and therefore idiosyncratic volatility, may be driven in part by differences in liquidity.

Durnev, et al. (2003, 2004) group each stock into industries based on the three-digit SIC industry grouping. This is done in order to reduce measurement error in the individual firm variables. In Appendix Tables A1 and A2, the results of Tables 1 and 2 are replicated, respectively using industry portfolio. Results, though somewhat less extreme, are remarkably similar. Industries with companies that have a worse information environment have lower average industry R^2.

3.2.4. Changes in the Information Environment

Previously we examined cross-sectional differences in the information environment and its association to differences in the level of market-model fit. In this section we examine if changes in the information environment as proxied by initiation of analyst coverage result in changes in R^2.

Using analyst forecasts from I/B/E/S we identify the first forecast ever made by any analyst or brokerage. Following the same procedure as in Table 1, we sort all stocks into R^2 portfolios based on the stock's R^2 in the year prior to the first analyst forecast. Table 3 reports the R^2 in the year prior to the first analyst forecast, the R^2 in the year following, the difference between the two, the bootstrapped standard error (using 1000 iterations) and the percentage of differences that are positive. Table 3 shows many significant differences pre and post initiation of analyst coverage. The differences are positive for low-R^2 stocks and negative for high-R^2 stocks, suggesting that differences in R^2s are a result of mean reversion rather than as a result of any improvement in the information environment that may have occurred when the first analyst began issuing forecasts. Nonetheless this table suggests that if initiation of analyst coverage signals an improvement in the information environment, then a better information environment is not associated with differences in R^2.

<INSERT TABLE 3 ABOUT HERE>

4. The Impact of Private Information on Market-model Fit

4.1. The Role of Private Information

The challenge with testing any measure of private information, such as R^2, is that private information is, of course, private. The best we can do is to look for other proxies for private information and see if there is a relation. To do this we look to the microstructure model of Easley, Kiefer and O'Hara (1997), hereafter EKO97. The model is both theoretically and empirically appealing for generating estimates of the probability of private information events. Its theoretical appeal comes from the fact that private information, at least that information which is profitable to the privately informed, can only be revealed to the market through trade. The EKO97 model conditions the market maker's beliefs about the probability of information events on the arrival of buy and sell trades – the very vehicle by which private information must be transmitted to the market. Empirically, its appeal comes from the fact that the EKO97 model is a purely trade based model. The measures derived from the EKO97 model do not use the same returns we use when calculating R^2. The measure of information based trading risk EKO97 derive (called PIN for the probability of information based trade), has similar characteristics to price/return based models of asymmetric information. In particular, high asymmetric information risk is associated with low trading volume and wide bid-ask spreads (Huang and Stoll, 1997 and Easley, Kiefer, O'Hara, and Paperman, 1996).

For our purposes the EKO97 model confers an additional advantage which return-based asymmetric information measures do not: as a byproduct of its estimation, the EKO97 model produces estimates of the probability of good and bad private information events. It is these estimates which we use to capture the average impact of private information on stock prices.

4.1.1. The Frequency of Private Information Events

If a low R^2 is a reflection of private information incorporation, then we would expect that low-R^2 stocks are those with relatively more private information events. The estimation of the probability of information events is described in Section 2.2. Table 4, Panel A presents the average probability of information events by NYSE-R^2 portfolio. The probability of information events is increasing in market-model R^2. A private information event is over a time and a half as likely for high-R^2 stocks than for low. Panel A also presents the average arrival rate of expected informed trades, uninformed trades and PIN by R^2 portfolio. The panels show that both informed trade (μ) and uninformed trade are increasing in R^2; however, uninformed trades are increasing more rapidly. This suggests that stocks with the greatest number private information events are also those with the greatest liquidity – a notion consistent with the models of Grossman (1976) and Kyle (1985), who posit that informed trade is profitable in expectation (and therefore undertaken), when there are liquidity traders among whose trades, the trades of the informed can be concealed. Panel B, confirms the associations suggested in the portfolio averages hold when examining correlations. In order to address the concern that measurement error has caused these associations, as robustness we follow Durnev, et al. (2003, 2004) and group in each stock into industries based on the three-digit SIC industry grouping. The results in Appendix Table A3, while weaker, confirm the individual stock level findings.

<INSERT TABLE 4 ABOUT HERE>

4.1.2. Calculating the Private Information Measure

We use the estimates of the arrival rate of informed and uninformed buys and sells, the probability of information events and the probability of good and bad news days from Eq. (2) to calculate the probability of good, bad, and no-information event days, conditional on the observed number of buys and sells for each trading day, following EKO97. For example, using Bayes rule, the probability of a good news event is:

$$\Pr(GoodNews|B,S) = \frac{\Pr(GoodNews)\times \Pr(B,S|GoodNews)}{\begin{pmatrix} \Pr(BadNews)\times \Pr(B,S|BadNews) \\ +\Pr(GoodNews)\times \Pr(B,S|GoodNews) \\ +\Pr(NoNews)\times \Pr(B,S|NoNews) \end{pmatrix}} \tag{3}$$

where B is the number of buys, S is the number of sales. Using the parameters from the EKO97 model in Eq. (2), Pr(*GoodNews*) is $\alpha(1\text{-}\delta)$, Pr(*BadNews*) is $\alpha\delta$, and Pr(*NoNews*) is $(1\text{-}\alpha)$. The probability of observing buys and sells are calculated in a manner similar to the maximum likelihood function in Eq. (2). For example, the ex-post probability of observing the number of buys and sells actually observed given that it was a good news day is:

$$\Pr(B,S \mid GoodNews) = \alpha(1-\delta)e^{-(\mu+\varepsilon_b)}\frac{(\mu+\varepsilon_b)^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!}, \tag{4}$$

where, as before, α is the unconditional probability of news occurring, $(1\text{-}\delta)$ is the probability of good news, conditional on there being news, μ is the arrival rate of informed trades conditional on there being news, and ε_b and ε_s are the arrival rates of uninformed trades.

Controlling for public news events, Roll (1988) finds low average market-model R^2s using return data at the daily and monthly frequency. He proposes that the low R^2s are the result of private information. In this section we directly address the central question of this paper: Does private information drive the low average R^2s of our pricing models? In order to address this question we examine the improvement in market-model R^2s when we include estimates of private information revelation. We begin this analysis by replicating Roll's (1988) results and controlling for known influences on the precision of model fit.

4.2. Idiosyncratic Volatility on News and Non-News Days

If private information were the primary driver of idiosyncratic returns and low R^2s, then we should see two things: first, days with news are more important for low-R^2 stocks than for high as seen in the magnitude of the idiosyncratic volatility on private-news days; second, because R^2s are measured annually, we should see a greater fraction of annual idiosyncratic volatility on private-news days than on non-news days. Ultimately, we find for the first, but not for the second. That is, there is greater idiosyncratic volatility on the average news day for low-R^2 stocks than for high, both relatively and in absolute terms. However, when we aggregate the idiosyncratic return up to the annual level, most of the idiosyncratic return occurs on non-news days. The fact that this is especially true for low-R^2 stock, suggests that R^2 is a poor proxy for firm-specific information incorporation as suggested by Morck, Yeung and Yu (2000, 2013), Durnev, Morck, Yeung and Zarowin (2003) and Durnev, Morck, and Yeung (2004).

We use the daily measure of the probability of a private information event described in Eq. (3) to decompose idiosyncratic volatility into the portion associated with private news days and that associated with Non-Private News days. Because the probability of a private information event measure is computed on a daily basis, instead of using weekly returns as in Tables 1-4, for these analyses we run the regression from Eq. (1) using daily returns. We obtain the residuals from the regression and calculate the portion of the Sum of Squared Errors (SSE) which occurs on days with a high probability of a private information event and those with a low probability of a private information event. In Panel A of Table 5, a day is considered an information event only if the probability of an information event was 90% or higher. In panel B we examine a much lower threshold, 50% and the results are very similar. As such, here we focus on the results in Panel A.

<INSERT TABLE 5 ABOUT HERE>

Consistent with the findings using annual data in Table 4, we see that low R^2-stocks have about half as many days with a high probability of an information event (columns 2 and 3). The middle three columns (columns 4-6) display the average amount of SSE that occurs on No News days, Good (private) News days, and Bad News days. Not surprisingly, the low-R^2 stocks have more idiosyncratic volatility over all. The interesting picture arises when examining the last three columns (columns 7-9), which shows the proportion of the total annual SSE that occurs on No News, Good News and Bad News days: about half the idiosyncratic volatility occurs on news days and the other on non-news days, and the differences between high- and low-R^2 portfolios are small. Another notable point is that because the roughly the same proportion of volatility is on News days vs. non-News days, whether a stock has a high R^2 or low, it suggests that each private-information-based trade has much greater impact on returns for low-R^2 stocks than on high. These findings make clear that differences in R^2 do not reflect differences in the amount of information conveyed into price: proportionately just as much private information is conveyed to low-R^2 stocks as high. In the next section we examine how private information impacts returns.

4.3. Private information, returns and model fit

4.3.1. Basic Model Fit

Next we examine how inclusion of private information improves R^2. To do this we will run regressions that include measures of private information calculated at the daily, weekly and monthly level. First, to see whether those measures of private information help explain returns we need to know what R^2s are without measures of private information. To do this we run the basic model from Eq. (1) at daily, weekly and monthly frequencies and get the average R^2 for each of the ten equally weighted NYSE-adjusted-R^2 portfolios and report them in Figure 1 (we exclude portfolios 5 and 6 to conserve space). The chart for each portfolio and frequency also shows 3 bars. These bars are the average portfolio adjusted R^2s for 3 different models.

The first bar in each chart is the average adjusted R^2 from the model in Eq. (1) except that here we regress individual stock returns on value-weighted market return and value-weighted two-digit SIC code industry returns for non-overlapping five-year windows from 1983 through 2002 instead of annual regressions using weekly data:

$$R_{i,t} = \alpha_i + \beta_{Mkt,i,t} R_{Mkt,t} + \beta_{Ind,i,t} R_{Ind_i \neq i,t} + \varepsilon_{i,t}. \tag{1'}$$

Bar 2 shows the average adjusted R^2 from a model that includes common sources of comovement, HML, SMB and WML and, at the monthly frequency only, we also include Pastor and Stambaugh's (2003) liquidity measure,[16]

$$\begin{aligned} R_{i,t} = \alpha_i + \sum_{n=-5}^{5} \beta_{Mkt,i,t+n} R_{Mkt,t+n} + \beta_{Ind,i,t} R_{Ind_i \neq i,t} \\ + \beta_{HML,i,t} HML_t + \beta_{SMB,i,t} SMB_t + \beta_{WML,i,t} WML_t + \varepsilon_{i,t}. \end{aligned} \tag{5}$$

<INSERT FIGURE 1 ABOUT HERE>

Comparing bars 1 and 2 across the portfolios displayed in Figure 1, we see that controlling for common sources of comovement improves adjusted R^2s, but not by large amount. For the entire sample adding HML, SMB and WML to (Eq. (5) improves model fit by 0.02, 0.03 and 0.04, for daily, weekly, and monthly frequencies. The fact that model fit improvement is increasing as the frequency decreases, suggests that measurement error also contributes to the poor explanatory power of the four-factor model; however, forming portfolios to reduce measurement error would impede our ability to examine the role of private firm-specific information in stock prices.

---

[16] HML, SMB and the momentum factor are from Kenneth French's Darmouth website. Pastor and Stambaugh liquidity measure is calculated for intra-month liquidity indicators, and as such cannot be calculated at daily frequencies. These data are provided courtesy of Lubos Pastor – thank you.

Bar 3 is based on the same regression as bar 2, but adds one lag of own returns to control for bid-ask bounce and leads and lags of the market to control for infrequent trading (Dimson, 1979).[17] Again, in the monthly regressions we include Pastor and Stambaugh's (2003) liquidity measure as well.

$$\begin{aligned} R_{i,t} &= \alpha_i + \sum_{n=-5}^{5} \beta_{Mkt,i,t+n} R_{Mkt,t+n} + \beta_{Ind_i i,t} R_{Ind_i \neq i,t} \\ &+ \beta_{HML,i,t} HML_t + \beta_{SMB,i,t} SMB_t + \beta_{WML,i,t} WML_t + \beta_{Own,i} R_{i,t-1} + \varepsilon_{i,t}. \end{aligned} \tag{6}$$

Comparing bars 2 and 3, we see that controls for infrequent trading and bid-ask bounce improve model fit only marginally, by 0.06 for the lowest daily R^2 portfolio and under 0.01 for higher R^2 portfolios. Since Table 1 shows us that low-R^2 stocks are illiquid and high-R^2 stocks are liquid, these findings are not surprising.

We will use Eq. (6) as our base-line model, when we include measures of private information in the next section. As such, in Table 6 we present detailed results for the adjusted R^2s from the third model in Eq. (6). We provide both full-sample adjusted-R^2 averages and results for NYSE only stock, because the private information measures used in the next section are calculated only for NYSE-listed stock. Differences between the entire sample and NYSE-sample-average R^2s are economically small.

Over all, controlling for missing factors and infrequent trading improves model fit, with the greatest improvement for the lowest R^2 stocks at the daily frequency. Nonetheless, for the lowest R^2 stocks over 90% of return remains unexplained. In the next section we examine what portion of return can be explained by private information.

[17] The choice of the number of leads and lags is arbitrary (as it was for Dimson, 1979). We did not experiment with different numbers of leads and lags.

4.3.2. Private Information and R^2

If private information causes the idiosyncratic volatility driving poor model fit, then including a proxy for private information should significantly improve the fit of the market model. In Table 7 we regress individual stock returns on the unconditional probability of a positive information event (good news) and the unconditional probability of a negative information event (bad news) controlling for value weighted market and industry returns, HML, SMB, and WML:

$$\begin{aligned} R_{i,t} &= \alpha_i + \sum_{n=-5}^{5} \beta_{Mkt,i,t+n} R_{Mkt,t+n} + \beta_{Ind_i i,t} R_{Ind_i \neq i,t} + \beta_{HML_i i,t} HML_t + \beta_{SMB_i i,t} SMB_t + \beta_{WML_i i,t} WML_t \\ &+ \beta_{Own,i} R_{i,t-1} + \beta_{Gi} \Pr(GoodNews \mid B,S) + \beta_{Bi} \Pr(BadNews \mid B,S) + \varepsilon_{i,t}, \end{aligned} \qquad (7)$$

where $R_{i,t}$ is the total return on individual stock i, $R_{Mkt,t}$ is the value-weighted market return, and $R_{Indi\neq i,t}$ is the value-weighted two-digit SIC industry return excluding firm i. HML and SMB are the Fama and French (1993) book-to-market and size factors, WML is the momentum factor, and $\Pr(GoodNews \mid B,S)$ and $\Pr(BadNews \mid B,S)$ are calculated as in Eq. (3). Five leads and lags of the market index are included to control for infrequent trading following Dimson (1979). There are two leads and lags at the weekly frequency and none at the monthly frequency. We also include one lag of own returns at the daily and weekly frequencies to control for bid-ask bounce. To aggregate the probabilities of good and bad news to the weekly and monthly level we simply calculate the sum of the probabilities over the period. In the monthly regressions we include Pastor and Stambaugh's (2003) liquidity measure as well.

Coefficients on the probabilities of good and bad news are what one would expect for a reasonable proxy for the evolution of information events: good news is associated with positive and significant returns and bad news, negative and significant. While information events are more prevalent for high-R^2 stocks as shown in Table 4, accounting for good and bad news events contributes more to the fit of low-R^2 portfolios, which we can see by comparing the results from Table 6, panel B to those in Table 7. For example, at the daily frequency, the average adjusted R^2

from Eq. (6) for low-R^2 stocks in portfolio 1 is 0.04 for ether the 1993-97 period or the 1998-02 period in Table 6, but accounting for private information, the adjusted R^2s jump to 0.15 in 1993-97 and 0.13 in 1998-02 in Table 7. By contrast, for high-R^2 stocks in portfolio 10, the average R^2 increases from 0.37 in Table 6 to 0.39 in Table 7 for the 1993-98 period and the adjusted R^2 actually decreases from 0.50 to 0.49 once we account for private information the 1998-02 period.

Roll (1988) found evidence that was consistent with notion that idiosyncratic volatility reflects private information; however, he conceded that his findings we also consistent with idiosyncratic volatility reflecting "a frenzy unrelated to concrete information." Since the information events are inferred from the level of trading activity, it is entirely possible that the probability of an information event is really just reflecting a high level of either buy or sell-side trading. This might be a particular concern in light of Duarte and Young's (2009) paper which suggests that the EKO97 PIN measure may merely proxy for illiquidity. In the next section, we further examine this possibility that "news" is really just "noise trading."

4.3.3. Private Information or Noise?

Because the private information measure is estimated from trade level data, it may be that the probability of good and bad news is really just a proxy for high transitory demand for liquidity or high demand- or supply-side pressure from noise traders. This is especially a concern because Duarte and Young (2009) provide evidence that PIN's strength in pricing assets comes in large part because it is highly correlated with illiquidity. To explore this possibility, we examine whether the returns associated with demand- and supply-side pressure are transitory or permanent. To do this we construct a measure of excess buy and sell-side pressure, which is the ratio of buys to sells. We subtract one from this ratio in order to normalize the measure to zero, so that a positive number indicates excess buy-side transactions and a negative number, excess sell-side transactions. We regress stock returns on our market model including the value-weighted market and industry returns, HML, SMB, WML, controls for bid-ask bounce and infrequent trading, and on contemporaneous and four lags of the buy/sell imbalance measures. We separate the imbalance measures into buy-side and sell-side, so as to allow for an asymmetric effect of buy and sell imbalance on returns. We perform the following regression to examine whether buy- and sells-side pressure is more indicative of information based trade causing permanent changes in return or non-information based trade causing temporary price impact:

$$\begin{aligned} R_{i,t} &= \alpha_i + \sum_{n=-5}^{5} \beta_{Mkt,i,t+n} R_{Mkt,t+n} + \beta_{Ind_i i,t} R_{Ind_i \neq i,t} \\ &+ \beta_{HML_i i,t} HML_t + \beta_{SMB_i i,t} SMB_t + \beta_{WML_i i,t} WML_t + \beta_{Own,i} R_{i,t-1} \\ &+ \sum_{i=0}^{4} \beta_{Buy,t-i} \left[\frac{B_{t-i}}{S_{t-i}} - 1 \middle| \frac{B_{t-i}}{S_{t-i}} - 1 \geq 0 \right] + \beta_{Sell,t-i} \left[\frac{B_{t-i}}{S_{t-i}} - 1 \middle| \frac{B_{t-i}}{S_{t-i}} - 1 < 0 \right] + \varepsilon_{i,t}, \end{aligned} \tag{8}$$

where all variables are defined as in Eq. (6). We focus on buys and sells as opposed to buy and sell-side volume, because Jones, Kaul and Lipson (1994) find that it is trades that are associated with volatility, volume plays a very minor role.

For the most part market the coefficients on the lagged trade imbalance measures are insignificant. However, negative coefficients are significant at the 5% level, between 10% and 20% of the time. This suggests that for this fraction of stocks a portion of the change in price due to sell side trading is partially reversed within four days. The results in Table 8 indicate that the effects of buy and sell imbalances have largely permanent effects on return, consistent with the notion that our private information measure is in fact associated with information and not noise trading. Taken together these findings suggest that private information does play a role in explaining poor model fit and high idiosyncratic volatility, however, it is also clear that private information only explains a fraction of returns. Roll (1988) finds that public information explains little of stock returns. Together the findings suggest that firm-specific information plays a relatively minor role in poor model fit and greater idiosyncratic volatility.

5. Conclusion

Morck, Yeung and Yu (2000) propose that R^2 or "synchronicity", as they call it, is an inverse measure "price informativeness", an intuitively appealing proposal in light of the fact that firm-specific information must negatively impact market-model R^2s. However, we have examined this proposition using several techniques, all of which lead to the same conclusion: a low market-model R^2 – high idiosyncratic volatility – is predominantly driven by factors other than private information.

If R^2 were inversely related to price informativeness, we would expect there to be more informed traders, and more likely than not, greater analyst coverage. We find the opposite. Low-R^2 stocks are covered by few analysts and have smaller increases in institutional ownership. This might not be a problem if arbitrageurs were able to profit from mispricings, thereby correcting prices. However, we find that low-R^2 stocks are smaller, making them less valuable to trade, and are less frequently traded, making them more difficult to trade, and have higher trading costs and price impact, making them less profitable to trade.

Using a microstructure model by Easley, Kiefer, and O'Hara (1997), which allows us to estimate the arrival of information on a daily basis, we have examined the effect of private information on prices and we have found that private information explains as much as 14% of returns for low-R^2 stocks regressions using weekly data. Nonetheless, for the same stocks over 80% of returns remain unexplained either by common sources of return comovement, or private information. In addition we show that high-R^2 stocks have many more frequent private information events than low-R^2 stocks. Overall, our evidence suggests that R^2 is a poor measure of information efficiency or stock price informativeness.

References

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31-56.

Atiase, Rowland K., 1985, Predisclosure information, firm capitalization, and security price behavior around earnings announcements, *Journal of Accounting Research*, 21-36.

Barclay, Michael J., Robert H. Litzenberger, and Jerold B. Warner, 1990, Private information, trading volume, and stock-return variances, *The Review of Financial Studies* 3, 233-253.

Bartram, Söhnke M., Gregory Brown and René M. Stulz, 2012, Why Are U.S. Stocks More Volatile? *The Journal of Finance* 67, 1329-1370.

Bessembinder, Hendrik, 2003, Quote-based competition and trade execution costs in NYSE-listed stocks, *Journal of Financial Economics* 70, 385-422.

Bessembinder, Hendrik and Subhrendu Rath, 2008, "Trading Costs and Return Volatility: Evidence From Exchange Listings" in *Market Liquidity*, edited by Greg N. Gregoriou and Francois-Serge Lhabitant, John Wiley & Sons.

Brennan, Michael J., Narasimhan Jegadeesh, and Bhaskaran Swaminathan, 1993, Investment analysis and the adjustment of stock prices to common information, *The Review of Financial Studies* 6, 799-824.

Bris, Arturo, William N. Goetzmann, and Ning Zhu, 2007, Efficiency and the bear: short sales and markets around the world, *Journal of Finance* 62(3), 1039-1079.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu, 2001, Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *The Journal of Finance* 56, 1-43.

Cao, Charles, Tim Simin and Jing Zhao, Can Growth Options Explain the Trend in Idiosyncratic Risk? *Review of Financial Studies*, 2008, 21, pp 2599-2633.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.

Chemmanur, Thomas J., and Paolo Fulghieri, 1999, A theory of the going-public decision, *Review of Financial Studies*, 249-279.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171-205.

Dasgupta, S., J. Gan, and N. Gao. 2010. Transparency, Price Informativeness, and Stock Return Synchronicity: Theory and Evidence. Journal of Financial and Quantitative Analysis 45 (5): 1189–1220.

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, Robert J. Waldmann, 1990, Noise trader risk in financial markets, *The Journal of Political Economy* 98, 703-738.

Dennis, Patrick and Deon Strickland, 2009, The Determinants of Idiosyncratic Volatility, University of Virginia working paper.

Diamond, Douglas W., and Robert E. Verrecchia, 1987, Constraints on short-selling and asset price adjustment to private information, *Journal of Financial Economics* 18, 277-311.

Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197-226.

Dow, James, Gary Gorton, 1994, Arbitrage chains, *Journal of Finance* 49, 819-849.

Duarte, J. and Young, L. 2009. Why is PIN priced? *Journal of Financial Economics*, 91(2):119–138.

Durnev, Art, Randall Morck, and Bernard Yeung, 2004, Value-enhancing capital budgeting and firm-specific stock return variation, *The Journal of Finance*, 59, 65-105.

Durnev, Art, Randall Morck, Bernard Yeung, and Paul Zarowin, 2003, Does greater firm-specific return variation mean more or less informed stock pricing? *Journal of Accounting Research* 41, 797-836.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is information risk a determinant of asset returns? *The Journal of Finance* 57, 2185-2221.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2010, Factoring Information Into Returns. *Journal of Financial and Quantitative Analysis* 45(2), 293-309.

Easley, David, Nicholas Kiefer, and Maureen O'Hara, 1997, One day in the life of a very common stock, *Review of Financial Studies* 10, 805-835.

Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman, 1996, Liquidity, information, and infrequently traded stocks, *The Journal of Finance* LI, 1405-1435.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Fink, Jason, Kristin Fink, Gustavo Grullon and James Weston (2010). What Drove the Increase in Idiosyncratic Volatility During the Internet Boom? *Journal of Financial and Quantitative Analysis*, Vol. 45, 1253-1278.

Frankel, Richard, and Xu Li, 2004, Characteristics of a Firm's Information Environment and the Information Asymmetry between Insiders and Outsiders, Working paper, MIT Sloan School of Management. *Journal of Accounting and Economics* 37, 229-259.

French, Kenneth R., and Richard Roll, 1986, Stock return variances: the arrival of information and the reaction of traders, *Journal of Financial Economics* 17, 5-26.

Gassen, Joachim, Ryan LaFond, Hollis A. Skaife and David Veenman, 2014, Illiquidity and stock price synchronicity. Working paper.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.

Gompers, Paul A., and Andrew Metrick, 1998, Institutional investors and equity prices, NBER working paper 6723.

Gompers, Paul A., and Andrew Metrick, 2001, Institutional investors and equity prices, *The Quarterly Journal of Economics*, 229-259.

Griffin, John M., and Michael Lemmon, 2002, Book-to-market equity, distress risk, and stock returns, *Journal of Finance*, 57, 2317–2336.

Grossman, Sanford J., 1976, On the efficiency of competitive stock markets where trades have diverse information, *The Journal of Finance* 31, 573-585

Grossman, Sanford J., Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393-408.

Grullon, Gustavo, George Kanatas, and James P. Weston, 2004, Advertising, Breadth of Ownership, and Liquidity, *The Review of Financial Studies* 17 (2), 439-461.

Heflin, Frank, Kenneth W. Shaw, and John J. Wild, 2005, Disclosure Policy and Market Liquidity: Impact of Depth Quotes and Order Sizes, *Contemporary Accounting Research* 22, 829-865.

Ho, Thomas S.Y., and Roni Michaely, 1988, Information quality of market efficiency, *The Journal of Financial and Quantitative Analysis* 23, 53-70.

Hong, Harrison, Terence Lim, and Jeremy C. Stein, 2000, Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies, *The Journal of Finance* 55, 265-295.

Hou, K., W. Xiong, and L. Peng. 2006. R2 and Price Inefficiency. Working paper.

Huang, Roger D., Hans R. Stoll, 1997, The Components of the Bid-Ask Spread: A General Approach, *The Review of Financial Studies* 10, 995-1034.

Irvine, P., and J. Pontiff, (2009). Idiosyncratic return volatility, cash flows and product market competition, *Review of Financial Studies*, 22(3), 1149-1177.

Jin, L., S.C. Myers, 2006, R-2 around the world: New theory and new tests, *Journal of Financial Economics,* 79 (2): 257-292.

Jones, Charles M., Gautam Kaul, and Marc L. Lipson, 1994. Information, trading, and volatility, *Journal of Financial Economics* 36, 127-154.

Jones, C. and P. Seguin, 1997, Transaction costs and price volatility: Evidence from commission deregulation, *American Economic Review*, 728-737.

Kim, Sok Tae, Ji-Chai Lin, and Myron B. Slovin, 1997, Market structure, informed trading, and analysts' recommendations, *The Journal of Financial and Quantitative Analysis* 32, 507-524.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1336.

Lee, D. W., and M. H. Liu. 2011. Does More Information in Stock Price Lead to Greater or Smaller Idiosyncratic Return Volatility? *Journal of Banking & Finance* 35 (6): 1563– 1580.

Lee, Charles M. C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 46, 733-746.

LeRoy, Stephen F., and Richard D. Porter, 1981, The present-value relation: tests based on implied variance bounds, *Econometrica* 49, 555-574.

Lesmond, David A., Joseph P. Ogden, and Charles A. Trzcinka, 1999, A new estimate of transaction costs, *The Review of Financial Studies* 12, 1113-1141.

Li, Kan, Randall Morck, Fan Yang, and Bernard Yeung, 2004, Firm-specific variation and openness in emerging markets, *Review of Economics and Statistics* 86, 658-69.

Li, B., S. Rajgopal, and M. Venkatachalam. 2013. R^2 and Idiosyncratic Risk are not Interchangeable. forthcoming in the *Accounting Review.*

Manning, Richard L., 1996, Logit regressions with continuous dependent variables measured with error, *Applied Economics Letters* 3, 183–184.

Merton, Robert C., 1987, A simple model of capital market equilibrium with incomplete information, *The Journal of Finance* 42, 483-510.

Morck, Randall, Bernard Yeung, and Wayne Yu, 2000, The information content of stock markets: why do emerging markets have synchronous stock price movements? *Journal of Financial Economics* 58, 215-260.

Morck, R., Yeung, B., Yu, W., 2013. R^2 and the Economy. Annual Review of Financial Economics 5, 143–166.

Pastor, Lubos, and Pietro Veronesi. 2003. Stock Valuation and Learning about Profitability. *Journal of Finance*, 58(5): 1749-1790.

Pastor, Lubos, and Robert, Stambaugh, 2003,Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.

Petersen, Mitchell A., 2009, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review Financial Studies* 22(1): 435-480.

Piotroski, Joseph D., and Darren T. Roulstone, 2004, The influence of analysts, institutional investors, and insiders on the incorporation market, industry, and firm-specific information into stock prices, *Accounting Review*, 79 1119-1152.

Roll, Richard, 1988, R^2, *The Journal of Finance*, 43, 541-566.

Rosett, R. N., 1959, A statistical model of friction in economics, *Econometrica* 27, 263-267.

Sapra, Sunil K., 1998, Bias and inefficiency of an ordinary least squares estimator for logit regressions with continuous dependent variables measured with error, *Applied Economics Letters* 5, 745–746.

Schwert, G. William, 1990, Stock Volatility and the Crash of '87, *The Review of Financial Studies* 3, 77-102.

Shiller, Robert J., 1981, The use of volatility measures in assessing market efficiency, *The Journal of Finance* 36, 291-304.

Umlauf, S., 1993, "Transactions Taxes and the Behavior of the Swedish Stock Market", *Journal of Financial Economics*, 33, 227-240.

West, Kenneth D., 1988, Dividend innovations and stock price volatility, *Econometrica* 56, 37-61.

Wei, S.X., C. Zhang, 2006, Why did individual stocks become more volatile? *Journal of Business*, 79, 259-292.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817-838.

Wurgler, Jeffrey, 2000, Financial markets and the allocation of capital, *Journal of Financial Economics* 58, 187-214.

Xing, X., and R. Anderson. 2011. Stock Price Synchronicity and Public Firm-Specific

Information. Journal of Financial Markets 14 (2): 259–276.

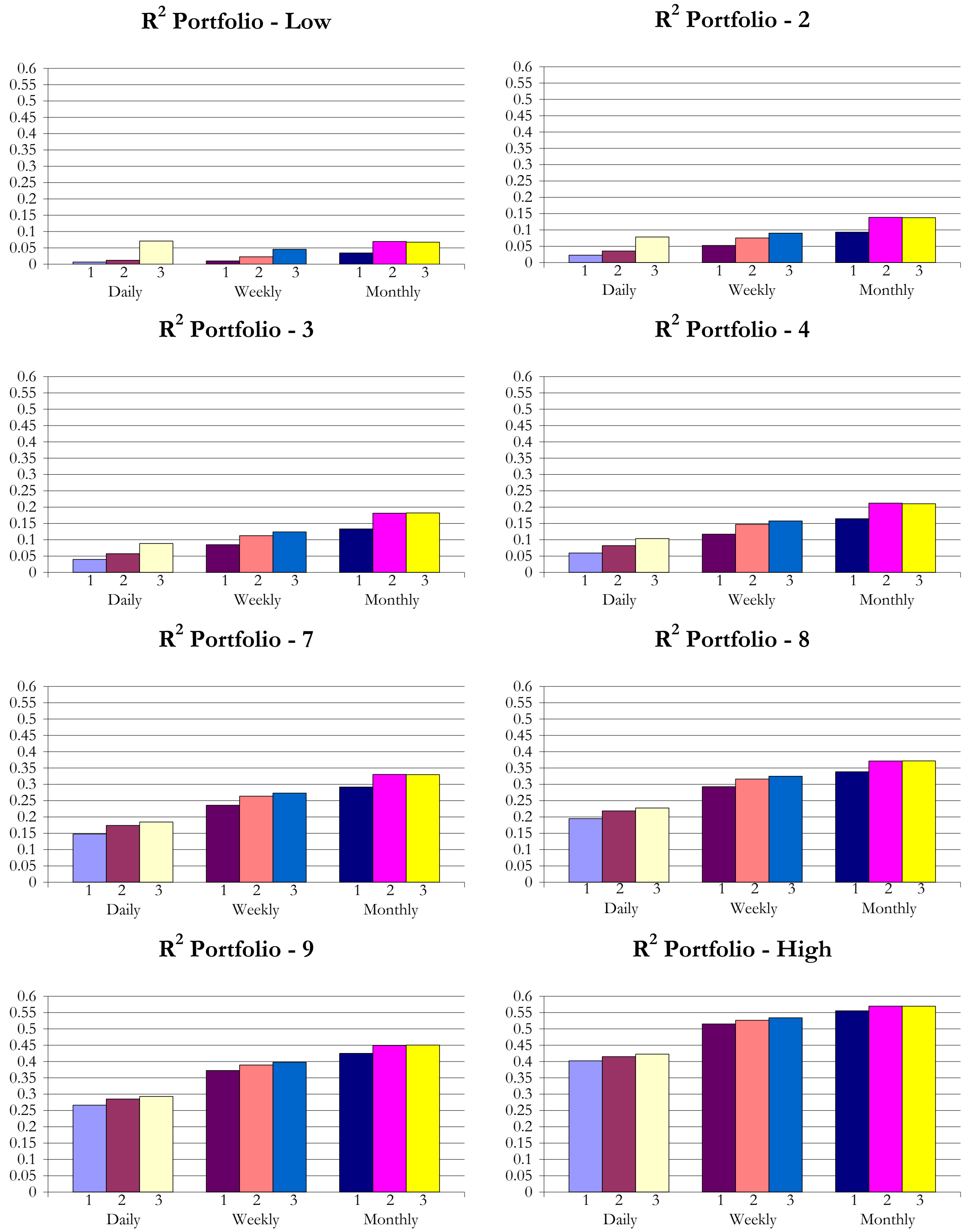

Figure 1. Market-model R²s. These figures display the average adjusted R²s for each of 10 NYSE-adjusted-R²-sorted portfolios for each of 3 models at 3 frequencies. The models regress total stock returns on

(1) the value-weighted market plus industries returns excluding the regressand,

(2) Model (1) plus the Carhart (1997) 4-factor model, and

(3) Model (2) with one lag of own returns to control for bid-ask bounce, and 5 leads and lags, 2 leads and lags and one lead and lag at the daily, weekly and monthly frequencies respectively.

Regressions are run in each of four 5-year non-overlapping windows from 1983 through 2002 and averaged over the entire period. Models (2) and (3) include the Pastor and Stambaugh (2003) illiquidity measure in regressions using monthly data. Data are common equity from CRSP. Portfolios 5 and 6 are excluded to conserve space.

Table 1. Summary Statistics for Market-Model R^2 Portfolios

To calculate R^2, the following regression is run for each stock in each year:

$$R_{i,t} = \alpha_i + \beta_{Mkt,i,t} R_{Mkt \neq i,t} + \beta_{Ind,i,t} R_{Ind_i \neq i,t}$$

where $R_{Mkt \neq,i,t}$ and $R_{indi \neq i,t}$ are the value weighted market return and company *i*'s two-digit SIC industry return excluding stock *i* in year *t*. Each year, common ordinary shares listed on NYSE, AMEX or NASDAQ with 52 weeks of weekly returns are sorted into ten NYSE-breakpoint portfolios based on R^2. For each data item from 1983 through 2002 averages are calculated for each R^2 portfolio, for each year *t*. Company counts are in December of year *t*. The average across all 20 years is presented in the table below. Size is the market capitalization at the end of December in year *t*-1. Age is the number of days listed on CRSP at the end of December in year *t*-1 divided by 365. Analyst Count is the number of unique analysts issuing forecasts in year *t*-1 as reported by IBES. Institutions and their holdings are counted following Gompers and Metrick (1998, 2001). The threshold is \$100 million in 1980 and is grown by the increase in the market capitalization of all common equity stock ever held by any institution. Change in breadth is the change in the number of unique institutions holding a stock from year *t*-2 to year *t*-1 divided by the number of institutions in year *t*-2 conditional on the firm being counted as an institutional holder of stock for any stock in both year *t*-2 and *t*-1 (i.e. new institutions are not counted as part of the change in breadth). Trading cost is a measure for year *t* of percentage trading costs, developed by Lesmond, Ogden and Trzcinka (1999); its construction is described in the text. Illiquidity in year *t* is calculated following Amihud (2002) and is the average in year *t* of the absolute value of the daily return divided by the daily dollar volume for all stocks with non-missing, non-zero volume. Illiquidity is multiplied by 10^6. Zero Vol Days (%) is the percentage of trading days with non-missing volume equal to zero in year *t*. In order to be included, all data must be available for each of the variables listed in this table. In panel C cross-sectional correlation coefficients are calculated each year from 1983 through 2002. The correlation coefficients are averaged across all 20 years. Below the diagonal are Pearson's correlation coefficients and above the diagonal are Spearman's rank correlation coefficients

Panel A

R^2 Portfolio Rank	Avg. R^2	Size (x 10^6)	Age	Analyst Count	Trading Cost (%)	Illiquidity (x 10^6)	Zero Vol Days (%)	Change in Breadth (%)
1	0.025	122	9	1	12.1	0.2114	18.7	0.3
2	0.072	233	10	2	8.1	0.1313	12.8	0.4
3	0.113	295	11	3	5.7	0.0779	9.1	0.6
4	0.153	496	12	5	3.8	0.0377	5.9	0.9
5	0.196	706	13	6	3.2	0.0297	4.2	1.1
6	0.244	1,155	14	8	2.0	0.0140	2.4	1.4
7	0.299	1,983	17	10	1.5	0.0055	1.3	1.8
8	0.364	2,942	19	13	1.0	0.0026	0.6	2.1
9	0.450	4,858	23	17	0.7	0.0006	0.2	2.6
10	0.595	7,479	29	23	0.4	0.0002	0.1	2.8
All Portfolios	0.151	1,020	13	5	6.7	0.1031	9.9	0.9

Table 1. (*continued*)

Panel B

R^2 Portfolio Rank	Counts NYSE	AMEX	NASDAQ	Average per Year	Total Observations
1	137	157	1052	1,346	26,910
2	137	96	604	837	16,730
3	137	71	432	639	12,786
4	136	51	315	502	10,038
5	136	39	238	413	8,265
6	137	28	179	344	6,876
7	136	16	132	284	5,675
8	137	12	92	241	4,815
9	137	5	56	198	3,958
10	135	4	22	161	3,215
Average per Year	1,364	478	3,121	4,963	
Total Observations	27,288	9,551	62,429		99,268

Panel C

	R^2	Size	Age	Analyst Count	Trading Cost	Illi-quidity	Zero Vol (%)	Change in Breadth
R^2		0.59	0.22	0.49	-0.60	-0.62	-0.52	0.26
Size	0.32		0.32	0.71	-0.86	-0.91	-0.72	0.36
Age	0.33	0.32		0.25	-0.36	-0.30	-0.18	-0.07
Analyst Count	0.58	0.51	0.41		-0.63	-0.69	-0.58	0.21
Trading Cost	-0.27	-0.09	-0.13	-0.23		0.86	0.63	-0.35
Illiquidity	-0.11	-0.03	-0.04	-0.10	0.43		0.80	-0.35
Zero Vol (%)	-0.33	-0.11	-0.13	-0.29	0.45	0.26		-0.26
Change in Breadth	0.27	0.12	-0.03	0.17	-0.14	-0.05	-0.15	

Table 2. Regressions of Market-model R^2 on Impediments to Informed Trade

This table presents the results of pooled time series/cross-sectional regressions of $ln(R^2/(1-R^2))$ on the impediments to trade measures defined in Table 1. To control for firm effects we use White (1980) heteroskedasticity corrected, (Rogers) clustered errors, clustering on firm. Coefficients on Age are multiplied by 1000. **, * indicate significance at the 1 and 5 percent level respectively.

	Panel A						
	Information Cost			Trade Cost	Illiquidity		Short Sale
	1	2	3	4	5	6	7
Intercept	-8.598 **	-2.573 **	-2.447 **	-2.011 **	-3.706 **	-1.916 **	-2.452 **
Size	0.355 **						
Age		0.077 **					
Analyst Count			0.308 **				
Trading Cost				-3.695 **			
Illiquidity					-0.265 **		
Zero Vol (%)						-3.179 **	
Change in Breadth							12.346 **
Adj. R-square	0.234	0.065	0.178	0.091	0.264	0.137	0.047
	Panel B						
							Robustness
	8	9	10	11	12	13	14
Intercept	-5.232 **	-8.431 **	-3.907 **	-6.947 **	-4.151 **	-4.036 **	-4.299 **
Size	0.105 **	0.329 **		0.249 **	0.029 **	0.031 **	0.028 **
Age		0.004 **		0.012 **	0.008 **	0.010 **	0.011 **
Analyst Count		0.093 **		0.105 **	0.080 **	0.084 **	0.084 **
Trading Cost			-0.294 **	0.126 *	-0.358 **	-0.044	-0.285 **
Illiquidity	-0.200 **		-0.265 **		-0.202 **	-0.183 **	-0.211 **
Zero Vol (%)				-1.024 **		-0.582 **	
Change in Breadth			1.920 **	3.978 **	2.833 **	3.068 **	2.694 **
Schwert Volatility							0.013 **
Adj. R-square	0.268	0.292	0.310	0.304	0.322	0.325	0.324

Table 3. Change in R^2 around Initiation of Analyst Coverage

R^2 and the R^2 portfolios are calculated as in Table 1. The table below reports the average R^2 for each R^2 portfolio and over the entire sample in the calendar year prior to the first earnings forecast reported by an analyst or brokerage through I/B/E/S and in the calendar year following. The difference between the R^2 following and prior to initiation of analyst coverage is also reported as is the *t*-test for the difference between the two. The *t*-test uses bootstrapped standard errors resampling 1,000 times with replacement. * indicates a significant differences at α=5%.

R^2 Portfolio Rank	Count	R^2 in Year Prior to Initiation of Analyst Coverage	R^2 in Year Following Initiation of Analyst Coverage	Difference	Bootstrap Standard Error	Difference Positive (%)
1	279	0.008	0.098	0.090*	0.006	89.2
2	288	0.026	0.105	0.079*	0.006	79.2
3	378	0.045	0.128	0.083*	0.006	73.0
4	356	0.071	0.120	0.049*	0.006	63.2
5	379	0.103	0.132	0.030*	0.006	50.4
6	386	0.134	0.164	0.030*	0.007	51.0
7	395	0.181	0.170	-0.012	0.007	40.0
8	375	0.236	0.196	-0.040*	0.007	36.5
9	379	0.310	0.254	-0.056*	0.008	33.5
10	350	0.458	0.355	-0.103*	0.009	26.3
All Stocks	3565	0.163	0.175	0.012*	0.003	52.7

Table 4. Asymmetric Information Risk, Private Information, Informed and Liquidity Trading and R^2

Each year, common ordinary shares listed on NYSE with available PIN measures are sorted into ten NYSE-breakpoint portfolios based on R^2, calculated as in Table 1. For each data item averages are calculated for each R^2 portfolio, for each year *t.* Using a model by Easley, Kiefer and O'Hara (1997), for each stock the following data items are estimated: Probability of an Information Event (α), Probability of Bad News (δ), Informed Trades (μ), Uninformed Buys(ε_b), Uninformed Sells (ε_s), and the probability of an informed trade (PIN). Panel A reports the the average of these for each of 10 NYSE-adjusted-R^2-sorted portfolios. The R^2s are calculated each year from a market-model regression of total stock returns on a value-weighted index and 2-digit SIC industry return (excluding the regressand firm) over 1993 through 2002 using weekly Wednesday-to-Wednesday returns. Data are for NYSE listed common equity only. Panel B reports Pearson correlation coefficients below the diagonal and Spearman rank correlation coefficients above.

Panel A

R^2 Portfolio Rank	Mean R^2	Pr. Info. Event (%) (α)	Pr. Bad News (%) (δ)	Expected Info. Trades ($\alpha\mu$)	Uninf. Buys (ε_b)	Uninf. Sells (ε_s)	PIN (%)
1	0.019	23.3	40.8	6	15	18	18.9
2	0.053	24.8	40.3	8	21	24	18.2
3	0.088	26.3	41.7	10	26	30	17.6
4	0.124	26.9	41.7	12	33	36	16.7
5	0.164	28.5	41.7	12	35	39	16.4
6	0.210	29.7	42.4	14	43	48	15.8
7	0.266	31.8	42.7	17	53	59	15.1
8	0.335	32.9	44.1	20	64	71	14.3
9	0.433	35.0	43.7	22	73	82	13.5
10	0.601	37.3	43.2	24	83	94	12.9
All Portfolios	0.229	29.6	42.2	14	45	50	15.9

Panel B

R^2 Portfolio Rank	R^2	Pr. Info. Event (%) (α)	Pr. Bad News (%) (δ)	Expected Info. Trades ($\alpha\mu$)	Uninf. Buys (ε_b)	Uninf. Sells (ε_s)	PIN (%)
R^2		0.39	0.04	0.44	0.47	0.46	-0.39
α	0.34		0.01	0.45	0.39	0.40	-0.06
δ	0.04	0.03		0.01	0.07	0.02	-0.14
$\alpha\mu$	0.36	0.34	-0.02		0.97	0.97	-0.39
ε_b	0.38	0.26	0.06	0.90		0.99	-0.56
ε_s	0.38	0.27	0.00	0.90	0.97		-0.55
PIN	-0.34	0.13	-0.12	-0.25	-0.42	-0.42	

Table 5. Private Information Events and Poor Model Fit

R^2 portfolios are calculated as in Table 4. For each day in the sample, the probability of a private information event is calculated following Eq. (3) using a model by Easley, Kiefer and O'Hara (1997). The top panel counts as news days all days with a probability of a private information event greater than .9. In panel B the probability must be greater than .5. The first three columns report the average percent of trading days with No News, Good News and Bad News. The second three columns report the cumulative total SSE occurring on the No News, Good News and Bad News days, were the SSE is from the annual regression using daily data described above. The last three columns displays the percent of the total SSE the SSE on No News, Good News and Bad News days represents. Each year the portfolio average for each portfolio is calculated. The figures below represent the average of these portfolio averages over 1993 through 2002.

Panel A: News is Probability of Information Event > .9

R^2 Portfolio	Days (%)			SSE occurring on			SSE as (%) of total occurring on		
Rank	No News	Good News	Bad News	No News	Good News	Bad News	No News	Good News	Bad News
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	84.6	9.5	5.9	0.541	0.273	0.151	55.8	28.4	15.8
2	82.6	10.9	6.6	0.518	0.272	0.154	54.6	29.0	16.4
3	81.2	11.5	7.3	0.505	0.264	0.158	54.1	28.7	17.2
4	80.0	12.2	7.8	0.491	0.264	0.158	53.4	29.1	17.5
5	79.0	12.9	8.1	0.472	0.253	0.161	52.8	28.9	18.3
6	77.2	13.7	9.0	0.448	0.262	0.158	51.1	30.5	18.4
7	75.4	15.0	9.7	0.419	0.248	0.160	50.2	30.3	19.5
8	73.8	15.1	11.1	0.391	0.227	0.167	49.3	29.3	21.4
9	71.5	16.9	11.6	0.354	0.209	0.145	49.3	30.2	20.5
10	70.1	18.3	11.6	0.294	0.156	0.109	51.7	29.1	19.2
All Portfolios	77.5	13.6	8.9	0.443	0.243	0.152	52.2	29.4	18.4

Panel B: News is Probability of Information Event > .5

R^2 Portfolio	Days (%)			SSE occurring on			SSE as (%) of total occurring on		
Rank	No News	Good News	Bad News	No News	Good News	Bad News	No News	Good News	Bad News
1	79.2	12.5	8.3	0.465	0.317	0.183	48.0	32.9	19.1
2	77.1	13.8	9.1	0.445	0.313	0.187	46.9	33.2	19.9
3	75.6	14.5	9.8	0.433	0.305	0.190	46.4	33.0	20.6
4	74.8	15.1	10.2	0.425	0.299	0.188	46.3	33.0	20.7
5	73.4	15.9	10.7	0.406	0.289	0.192	45.5	32.8	21.8
6	71.9	16.5	11.6	0.386	0.294	0.187	44.2	34.1	21.7
7	70.2	17.7	12.2	0.363	0.277	0.187	43.6	33.7	22.7
8	68.7	17.7	13.5	0.341	0.253	0.192	43.0	32.5	24.5
9	66.3	19.6	14.2	0.308	0.232	0.168	43.1	33.4	23.5
10	65.0	20.9	14.1	0.258	0.174	0.127	45.8	32.0	22.2
All Portfolios	72.2	16.4	11.4	0.383	0.275	0.180	45.3	33.1	21.7

Table 6. Average Portfolio R^2s for Carhart Four-Factor Model Controlling for Bid-Ask Bounce and Infrequent Trading

The table reports the average adjusted R^2s for each of 10 NYSE-adjusted-R^2-sorted portfolios from the following:

$$R_{i,t} = \alpha_i + \sum_{n=-5}^{5} \beta_{Mkt,i,t+n} R_{Mkt,t+n} + \beta_{Ind_i,i,t} R_{Ind_i \neq i,t} + \beta_{HML,i,t} HML_t + \beta_{SMB,i,t} SMB_t + \beta_{WML,i,t} WML_t + \beta_{Own,i} R_{i,t-1} + \varepsilon_{i,t}, \tag{6'}$$

where $R_{i,t}$ is the total return on individual stock i, $R_{Mkt,t}$ is the value-weighted market return, and $R_{Indi \neq i,t}$ is the value-weighted two-digit SIC industry return excluding firm i. HML, SMB and WML are from Ken French's website. Five leads and lags of the market index control for infrequent trading at the daily frequency, but only two leads and lags at the weekly frequency and none at monthly. One lag of own returns is included at the daily and weekly frequencies to control for bid-ask bounce. At the monthly frequency Pastor and Stambaugh's (2003) liquidity factor is included. Regressions are run in each of four 5-year non-overlapping windows from 1983 through 2002 and averaged over the entire period. Data are common equity from CRSP. Averages over portfolios and years are the equally weighted average of the portfolio averages.

Portfolios		1	2	3	4	5	6	7	8	9	10	Average
Panel A: Entire Market												
Daily:	83-87	0.06	0.10	0.13	0.16	0.18	0.22	0.25	0.28	0.34	0.45	0.21
	88-92	0.07	0.08	0.08	0.09	0.11	0.13	0.17	0.21	0.27	0.38	0.16
	93-97	0.09	0.07	0.07	0.07	0.08	0.09	0.12	0.16	0.22	0.36	0.13
	98-02	0.06	0.06	0.08	0.10	0.13	0.15	0.21	0.26	0.34	0.50	0.19
Weekly:	83-87	0.05	0.14	0.19	0.23	0.27	0.31	0.35	0.39	0.45	0.55	0.29
	88-92	0.04	0.08	0.12	0.16	0.19	0.23	0.28	0.33	0.41	0.53	0.24
	93-97	0.04	0.05	0.07	0.09	0.12	0.15	0.19	0.25	0.33	0.49	0.18
	98-02	0.05	0.08	0.11	0.15	0.18	0.22	0.27	0.33	0.41	0.57	0.24
Monthly:	83-87	0.09	0.22	0.26	0.31	0.36	0.42	0.44	0.49	0.55	0.64	0.38
	88-92	0.07	0.13	0.19	0.21	0.25	0.28	0.33	0.38	0.46	0.58	0.29
	93-97	0.04	0.07	0.11	0.12	0.15	0.18	0.22	0.25	0.34	0.49	0.20
	98-02	0.08	0.13	0.17	0.20	0.24	0.27	0.32	0.36	0.45	0.57	0.28
Panel B: NYSE Only												
Daily:	83-87	0.07	0.12	0.16	0.19	0.21	0.24	0.27	0.30	0.35	0.45	0.24
	88-92	0.04	0.05	0.07	0.07	0.11	0.13	0.18	0.22	0.28	0.38	0.15
	93-97	0.04	0.06	0.07	0.08	0.09	0.11	0.13	0.17	0.23	0.37	0.13
	98-02	0.04	0.06	0.09	0.13	0.16	0.18	0.23	0.27	0.35	0.50	0.20
Weekly:	83-87	0.07	0.14	0.20	0.24	0.29	0.32	0.35	0.40	0.45	0.55	0.30
	88-92	0.05	0.09	0.12	0.16	0.18	0.23	0.29	0.33	0.40	0.53	0.24
	93-97	0.04	0.05	0.08	0.10	0.13	0.15	0.19	0.24	0.33	0.49	0.18
	98-02	0.05	0.09	0.12	0.15	0.18	0.23	0.28	0.33	0.41	0.58	0.24
Monthly:	83-87	0.11	0.21	0.27	0.32	0.37	0.43	0.44	0.49	0.55	0.65	0.38
	88-92	0.13	0.15	0.17	0.23	0.26	0.29	0.33	0.40	0.47	0.57	0.30
	93-97	0.04	0.08	0.13	0.10	0.13	0.17	0.19	0.24	0.34	0.50	0.19
	98-02	0.07	0.14	0.15	0.18	0.21	0.24	0.31	0.35	0.43	0.56	0.26
Panel C: Portfolio Average												
Entire Market:												
Daily		0.07	0.08	0.09	0.10	0.12	0.15	0.18	0.23	0.29	0.42	0.17
Weekly		0.05	0.09	0.12	0.16	0.19	0.23	0.27	0.32	0.40	0.53	0.24
Monthly		0.07	0.14	0.18	0.21	0.25	0.29	0.33	0.37	0.45	0.57	0.29
NYSE Only:												
Daily		0.05	0.07	0.10	0.12	0.14	0.17	0.20	0.24	0.30	0.43	0.18
Weekly		0.05	0.09	0.13	0.16	0.20	0.23	0.28	0.33	0.40	0.54	0.24
Monthly		0.09	0.14	0.18	0.21	0.24	0.28	0.32	0.37	0.45	0.57	0.28

Table 7. Impact of Good and Bad Private Information Events on Returns

From the following regression this table reports the average adjusted R^2s, average coefficients on the probability of good and bad news events, and the percent of coefficients which are significant at the 5% level for each of 10 NYSE Adjusted R^2 sorted portfolios:

$$R_{i,t} = \alpha_i + \sum_{n=-5}^{5} \beta_{Mkt,i,t+n} R_{Mkt,t+n} + \beta_{Ind_i i,t} R_{Ind_i \neq i,t} + \beta_{HML_i i,t} HML_t + \beta_{SMB_i i,t} SMB_t + \beta_{WML_i i,t} WML_t \qquad (7')$$
$$+ \beta_{Own,i} R_{i,t-1} + \beta_{Gi} \Pr(GoodNews \mid B,S) + \beta_{Bi} \Pr(BadNews \mid B,S) + \varepsilon_{i,t},$$

All variables are defines as in Table 6. Estimates of good and bad news events are from Eq. (3) as described in the text. At the monthly frequency Pastor and Stambaugh's (2003) liquidity factor is included. Regressions are run in each of four 5-year non-overlapping windows from 1983 through 2002 and averaged over the entire period. Coefficients are multiplied by 100.

Portfolios		1	2	3	4	5	6	7	8	9	10	Average
						Panel A: Daily						
R2	93-97	0.15	0.16	0.16	0.17	0.18	0.19	0.20	0.23	0.27	0.39	0.21
	98-02	0.13	0.15	0.16	0.19	0.21	0.23	0.28	0.30	0.36	0.49	0.25
Beta	93-97	3.48	2.69	2.26	2.07	1.70	1.57	1.31	1.13	0.97	0.58	1.78
Good	98-02	3.34	2.76	2.00	1.88	1.68	1.51	1.34	1.14	1.01	0.75	1.74
Beta	93-97	-2.35	-1.40	-1.21	-1.01	-0.83	-0.59	-0.48	-0.39	-0.26	-0.15	-0.87
Bad	98-02	-2.16	-1.88	-1.50	-1.33	-1.21	-1.00	-0.95	-0.75	-0.54	-0.42	-1.18
Pct Sig	93-97	98	99	98	100	100	99	99	99	95	92	98.0
Good	98-02	98	98	97	97	99	100	96	95	94	93	96.7
Pct Sig	93-97	78	79	72	77	72	65	60	57	47	43	64.9
Bad	98-02	79	82	85	79	82	78	79	73	59	54	75.1
						Panel B: Weekly						
R2	93-97	0.18	0.17	0.19	0.20	0.20	0.23	0.25	0.29	0.35	0.50	0.26
	98-02	0.16	0.18	0.18	0.21	0.24	0.28	0.32	0.36	0.43	0.58	0.29
Beta	93-97	2.69	2.25	1.74	1.53	1.20	1.10	0.79	0.66	0.52	0.25	1.27
Good	98-02	2.71	2.29	1.45	1.34	1.28	1.04	0.92	0.70	0.61	0.44	1.28
Beta	93-97	-1.44	-0.83	-0.69	-0.59	-0.44	-0.24	-0.20	-0.15	-0.06	-0.02	-0.47
Bad	98-02	-1.68	-1.47	-1.15	-0.95	-0.99	-0.84	-0.84	-0.69	-0.51	-0.40	-0.95
Pct Sig	93-97	85	89	89	89	84	81	69	64	59	40	74.9
Good	98-02	79	86	72	78	74	78	68	63	54	48	70.0
Pct Sig	93-97	45	39	35	37	32	23	23	22	19	16	29.1
Bad	98-02	47	58	49	48	44	47	53	42	43	38	46.8
						Panel B: Monthly						
R2	93-97	0.17	0.16	0.21	0.19	0.18	0.23	0.24	0.27	0.35	0.50	0.25
	98-02	0.17	0.21	0.21	0.22	0.25	0.26	0.34	0.37	0.46	0.58	0.31
Beta	93-97	1.78	1.24	0.87	0.89	0.60	0.49	0.34	0.20	0.16	0.04	0.66
Good	98-02	1.77	1.50	0.91	0.63	0.75	0.63	0.50	0.36	0.37	0.21	0.76
Beta	93-97	-0.66	-0.45	-0.19	-0.32	-0.15	-0.02	-0.03	-0.05	0.01	-0.01	-0.19
Bad	98-02	-1.11	-1.19	-0.77	-0.56	-0.82	-0.68	-0.60	-0.70	-0.45	-0.24	-0.71
Pct Sig	93-97	45	38	36	41	26	24	20	12	11	8	26.1
Good	98-02	32	37	25	22	19	22	23	18	17	14	22.9
Pct Sig	93-97	15	12	11	15	7	8	8	6	6	6	9.3
Bad	98-02	14	21	19	18	18	16	14	18	18	9	16.5

Table 8. Impact of Buy/Sell Imbalance on Daily Returns

From the following regression this table reports the average adjusted R^2s, average coefficients times 100 on the buy/sell imbalance measures as ((# of buys/# of sells) -1), and the percent of coefficients which are significant at the 5% level for each of 10 NYSE-adjusted-R^2-sorted portfolios.

$$
\begin{aligned}
R_{i,t} &= \alpha_i + \sum_{n=-5}^{5} \beta_{Mkt,i,t+n} R_{Mkt,t+n} + \beta_{Ind_i i,t} R_{Ind_i \neq i,t} \\
&\quad + \beta_{HML_i i,t} HML_t + \beta_{SMB_i i,t} SMB_t + \beta_{WML_i i,t} WML_t + \beta_{Own,i} R_{i,t-1} \\
&\quad + \sum_{i=0}^{4} \beta_{Buy,t-i} \left[\frac{B_{t-i}}{S_{t-i}} - 1 \middle| \frac{B_{t-i}}{S_{t-i}} - 1 \geq 0 \right] + \beta_{Sell,t-i} \left[\frac{B_{t-i}}{S_{t-i}} - 1 \middle| \frac{B_{t-i}}{S_{t-i}} - 1 < 0 \right] + \varepsilon_{i,t},
\end{aligned}
\qquad (8')
$$

Each stock's total return is regressed on value-weighted market plus industries returns excluding the regressand, plus HML, SMB and WML with one lag of own returns to control for bid-ask bounce, and 5 leads and lags, 2 leads and lags and one lead and lag at the daily, weekly and monthly frequencies respectively to control for infrequent trading, and the probability of good and bad news events. Regressions are run in each of four 5-year non-overlapping windows from 1983 through 2002 and averaged over the entire period. The percentage of stocks significant at the 95% is also indicated below the coefficients.

Portfolios	c	1	2	3	4	5	6	7	8	9	10	Average
Panel A: R^2												
R2	93-97	0.14	0.16	0.17	0.17	0.19	0.19	0.20	0.23	0.28	0.40	0.21
	98-02	0.12	0.15	0.17	0.19	0.21	0.23	0.28	0.30	0.37	0.51	0.25
Panel B: Contemporaneous												
Beta	93-97	0.42	0.34	0.30	0.30	0.29	0.27	0.27	0.33	0.34	0.23	0.31
Pos Imb	98-02	0.58	0.56	0.70	0.79	0.77	0.98	1.12	1.15	1.07	1.10	0.88
Pct	93-97	72	77	71	71	69	51	55	49	48	38	60.0
Pos Sig	98-02	63	76	75	78	80	84	87	83	76	75	77.9
Pct	93-97	0	0	0	0	1	1	0	0	0	2	0.3
Neg Sig	98-02	0	0	0	0	0	0	0	0	1	1	0.1
Beta	93-97	2.51	2.21	2.11	2.01	1.97	1.96	1.80	1.75	1.51	1.03	1.89
Neg Imb	98-02	3.20	3.09	2.86	3.09	3.06	3.01	2.80	2.90	2.50	2.00	2.85
Pct	93-97	99	98	99	98	99	99	98	96	94	88	96.8
Pos Sig	98-02	97	99	96	98	96	98	93	95	91	85	94.7
Pct	93-97	0	0	0	1	0	0	0	0	1	0	0.1
Neg Sig	98-02	0	0	0	0	0	0	1	0	1	1	0.2
Panel C: Lag 1												
Beta	93-97	0.00	0.00	-0.02	0.00	-0.03	-0.05	-0.05	-0.02	-0.05	-0.02	-0.02
Pos Imb	98-02	-0.01	-0.04	-0.12	-0.17	-0.14	-0.18	-0.18	-0.20	-0.19	-0.21	-0.14
Pct	93-97	5	5	4	1	3	1	2	4	3	3	3.1
Pos Sig	98-02	2	4	1	1	0	2	0	1	0	1	1.2
Pct	93-97	5	6	2	3	6	6	11	4	4	5	5.1
Neg Sig	98-02	3	8	11	10	8	9	12	10	8	11	9.0
Beta	93-97	-0.24	-0.20	-0.24	-0.22	-0.20	-0.16	-0.16	-0.16	-0.10	-0.03	-0.17
Neg Imb	98-02	-0.49	-0.45	-0.51	-0.61	-0.52	-0.57	-0.42	-0.52	-0.52	-0.29	-0.49
Pct	93-97	1	2	0	0	0	1	1	1	2	3	1.0
Pos Sig	98-02	0	1	0	1	1	0	1	0	0	1	0.5
Pct	93-97	17	17	11	15	14	12	13	10	8	4	12.1
Neg Sig	98-02	16	15	14	25	19	20	13	18	16	11	16.8

Table 8. Impact of Buy/Sell Imbalance on Daily Returns *(continued)*

Panel D: Lag 2												
Beta	93-97	-0.01	-0.02	-0.01	-0.01	0.00	-0.02	-0.01	-0.02	-0.03	-0.04	-0.02
Pos Imb	98-02	-0.06	-0.03	-0.10	-0.07	-0.12	-0.11	-0.16	-0.16	-0.06	-0.14	-0.10
Pct	93-97	4	2	3	1	3	3	2	2	2	2	2.3
Pos Sig	98-02	3	3	4	1	0	1	4	0	1	1	1.8
Pct	93-97	6	4	5	2	5	6	4	1	4	3	3.9
Neg Sig	98-02	1	2	6	5	7	8	6	5	5	6	5.0
Beta	93-97	-0.21	-0.11	-0.21	-0.17	-0.24	-0.18	-0.19	-0.22	-0.19	-0.11	-0.18
Neg Imb	98-02	-0.29	-0.30	-0.28	-0.44	-0.33	-0.40	-0.42	-0.49	-0.47	-0.43	-0.39
Pct	93-97	0	0	1	1	0	2	0	1	0	1	0.5
Pos Sig	98-02	0	1	0	0	0	0	0	1	0	0	0.3
Pct	93-97	10	7	14	7	18	9	14	16	12	7	11.5
Neg Sig	98-02	8	12	6	11	7	12	15	15	15	13	11.4
Panel E: Lag 3												
Beta	93-97	-0.01	-0.01	0.00	0.01	-0.02	0.00	-0.01	-0.03	0.01	0.02	0.00
Pos Imb	98-02	-0.02	-0.01	-0.02	-0.03	-0.06	-0.05	-0.06	0.02	-0.10	-0.12	-0.04
Pct	93-97	6	1	2	4	1	1	1	1	4	2	2.3
Pos Sig	98-02	2	1	4	2	1	1	3	2	1	1	1.8
Pct	93-97	1	4	3	1	4	5	5	6	2	3	3.3
Neg Sig	98-02	3	3	4	2	5	5	3	4	5	7	4.1
Beta	93-97	-0.23	-0.17	-0.20	-0.17	-0.14	-0.19	-0.16	-0.16	-0.16	-0.15	-0.17
Neg Imb	98-02	-0.31	-0.21	-0.22	-0.28	-0.31	-0.32	-0.30	-0.42	-0.28	-0.23	-0.29
Pct	93-97	0	1	0	1	1	0	1	1	1	0	0.4
Pos Sig	98-02	0	1	2	1	1	1	1	1	1	1	1.0
Pct	93-97	9	7	12	8	7	12	9	10	7	12	9.4
Neg Sig	98-02	9	7	8	11	8	12	8	13	11	9	9.5
Panel F: Lag 4												
Beta	93-97	-0.03	-0.02	-0.02	-0.01	0.00	-0.02	-0.01	0.00	-0.03	0.00	-0.01
Pos Imb	98-02	0.00	-0.04	-0.04	-0.06	-0.07	-0.09	-0.15	-0.21	-0.06	-0.12	-0.08
Pct	93-97	6	3	2	2	3	2	3	2	2	4	2.8
Pos Sig	98-02	3	3	2	1	1	2	0	0	2	1	1.4
Pct	93-97	6	1	3	3	3	3	3	2	3	3	3.0
Neg Sig	98-02	3	4	4	4	7	5	7	5	3	4	4.6
Beta	93-97	-0.19	-0.15	-0.17	-0.16	-0.15	-0.16	-0.18	-0.19	-0.16	-0.13	-0.16
Neg Imb	98-02	-0.30	-0.19	-0.26	-0.24	-0.35	-0.31	-0.30	-0.31	-0.37	-0.22	-0.29
Pct	93-97	0	1	0	1	0	1	1	0	1	1	0.5
Pos Sig	98-02	1	1	0	1	1	1	1	0	1	2	0.7
Pct	93-97	5	7	11	10	9	8	9	12	8	13	9.1
Neg Sig	98-02	8	5	4	10	8	9	11	7	13	6	8.1

Appendix Table A1. Summary Statistics for Industry Market-Model R^2 Portfolios

All measures are calculated as in Table 1, except that stocks are first grouped by their 3-digit SIC industry. The 3-digit industry-average R^2s are used to sort industries into an R^2 portfolio. The results below are the R^2-portfolio average of the industry-average measures.

Panel A								
					Trading		Zero	Change in
R^2 Portfolio	Avg.	Size		Analyst	Cost	Illiquidity	Vol Days	Breadth
Rank	R^2	(x 10^8)	Age	Count	(%)	(x 10^6)	(%)	(%)
1	0.053	201	12	2	11.9	0.3171	17.5	0.4
2	0.083	363	12	3	9.3	0.1300	13.7	0.5
3	0.100	453	12	3	8.1	0.1422	12.3	0.6
4	0.113	518	13	4	7.6	0.1132	11.3	0.7
5	0.126	636	13	4	6.9	0.1117	11.3	0.8
6	0.140	780	14	4	6.3	0.0840	9.8	0.8
7	0.156	870	15	5	5.7	0.0878	8.7	0.9
8	0.176	1,249	15	6	5.2	0.0608	7.9	1.1
9	0.208	1,516	16	7	4.6	0.0751	6.6	1.3
10	0.303	3,386	22	11	3.7	0.0639	4.3	1.4
All Portfolios	0.146	994	14	5	6.9	0.1185	10.4	0.9

Appendix Table A1. (*continued*)

Panel B

	R^2	Size	Age	Analyst Count	Trading Cost	Illi-quidity	Zero Vol (%)	Change in Breadth
R^2		0.59	0.27	0.58	-0.45	-0.36	-0.42	0.27
Size	0.45		0.41	0.77	-0.45	-0.32	-0.40	0.22
Age	0.34	0.40		0.33	-0.36	-0.23	-0.16	0.05
Analyst Count	0.64	0.59	0.44		-0.45	-0.34	-0.46	0.20
Trading Cost	-0.34	-0.18	-0.26	-0.35		0.76	0.60	-0.21
Illiquidity	-0.18	-0.10	-0.10	-0.17	0.53		0.62	-0.18
Zero Vol (%)	-0.40	-0.22	-0.19	-0.42	0.51	0.32		-0.15
Change in Breadth	0.26	0.12	0.07	0.20	-0.15	-0.11	-0.14	

Appendix Table A2. Regressions of Industry Market-model R^2 on Impediments to Informed Trade

As in Appendix Table A1, all measures are calculated as in Table 2, except that before aggregating into R^2 portfolio, the 3-digit SIC industry average is calculated for each measure.

Panel A														
			Information Cost				Trade Cost			Illiquidity			Short Sale	
	1		2		3		4		5		6		9	
Intercept	-4.510	**	-2.058	**	-1.996	**	-1.690	**	-2.315	**	-1.570	**	-2.026	**
Size	0.142	**												
Age			0.000	**										
Analyst Count					0.362	**								
Trading Cost							-1.894	**						
Illiquidity									-0.138	**				
Zero Vol (%)											-2.839	**		
Change in Breadth													12.256	**
Adj. R-square	0.095		0.049		0.226		0.042		0.110		0.126		0.048	
Panel B														
													Robustness	
	10		11		12		13		14		15		16	
Intercept	-4.193	**	-4.205	**	-2.387	**	-3.443	**	-4.065	**	-3.723	**	-3.955	**
Size	0.104	**	0.113	**			0.072	**	0.092	**	0.079	**	0.097	**
Age			0.000				0.000		0.000		0.000		0.000	
Analyst Count			0.253	**			0.241	**	0.231	**	0.226	**	0.222	**
Trading Cost					-0.103		0.741	**	0.534	**	0.910	**	0.389	
Illiquidity	-0.110	**			-0.115	**			-0.049	**	-0.036	**	-0.036	**
Zero Vol (%)							-1.058	**			-0.822	**		
Change in Breadth					8.520	**	6.917	**	6.220	**	6.231	**	6.229	**
Schwert Volatility													-0.022	**
Adj. R-square	0.156		0.258		0.120		0.282		0.281		0.286		0.284	

Appendix Table A3. Industry Average Asymmetric Information Risk, Private Information, Informed and Liquidity Trading and R^2

All measures are calculated as in Table 4, except that before aggregating into R^2 portfolio, the 3-digit SIC industry average is calculated for each measure.

Panel A

R^2 Portfolio Rank	Mean R^2	Pr. Info. Event (%) (α)	Pr. Bad News (%) (δ)	Expected Info. Trades ($\alpha\mu$)	Uninf. Buys (ε_b)	Uninf. Sells (ε_s)	PIN (%)
1	0.055	25.1	39.0	10	26	29	17.5
2	0.077	26.6	40.7	10	27	31	17.6
3	0.090	26.9	43.3	11	34	38	16.5
4	0.102	27.3	42.4	12	36	40	16.6
5	0.115	28.2	42.7	12	36	40	16.7
6	0.128	29.1	41.6	12	36	40	16.7
7	0.144	28.7	41.3	15	46	50	15.9
8	0.165	30.2	43.1	16	49	54	15.8
9	0.200	30.4	42.1	17	55	61	15.6
10	0.299	32.2	43.1	18	60	66	15.0
All Portfolios	0.138	28.5	41.9	13	41	45	16.4

Panel B

R^2 Portfolio Rank	R^2	Pr. Info. Event (%) (α)	Pr. Bad News (%) (δ)	Expected Info. Trades ($\alpha\mu$)	Uninf. Buys (ε_b)	Uninf. Sells (ε_s)	PIN (%)
R^2		0.32	0.06	0.40	0.42	0.41	-0.22
α	0.27		0.08	0.43	0.39	0.41	-0.07
δ	0.05	0.07		0.02	0.09	0.04	-0.12
$\alpha\mu$	0.29	0.35	-0.01		0.96	0.96	-0.39
ε_b	0.30	0.29	0.08	0.92		0.99	-0.53
ε_s	0.31	0.30	0.03	0.93	0.98		-0.53
PIN	-0.22	0.08	-0.10	-0.32	-0.46	-0.47	