What moves stock prices?
The role of news, noise, and information☆

Jonathan Brogaard [a], Huong Nguyen [b], Talis J. Putnins [b, c], Eliza Wu [d]

[a] *University of Utah*
[b] *University of Technology Sydney*
[c] *Stockholm School of Economics in Riga*
[d] *University of Sydney*

Review of Financial Studies, Forthcoming

October 2021

Abstract

We develop a return variance decomposition model to separate the role of different types of information and noise in stock price movements. We disentangle four components: market-wide information, private firm-specific information revealed through trading, firm-specific information revealed through public sources, and noise. Overall, 31% of the return variance is from noise, 37% from public firm-specific information, 24% from private firm-specific information, and 8% from market-wide information. Since the mid-1990s, there has been a dramatic decline in noise and an increase in firm-specific information, consistent with increasing market efficiency.

JEL classification: G12; G14; G15

Keywords: variance decomposition; firm-specific information; market-wide information; stock return synchronicity

☆ The internet appendix that accompanies this paper can be obtained here: https://bit.ly/3FcV9UR
We thank the editor (Itay Goldstein), two anonymous referees, Paul Besson, Jean-Philippe Bouchaurd, Amit Goyal, Joel Hasbrouck, Nikolaus Hautsch, Petri Jylhä, Alok Kumar, Charles-Albert Lehalle, Joshua Mollner, Ryan Riordan, Elvira Sojli, and seminar and conference participants at the American Finance Association Annual Meeting, the European Finance Association Annual Meeting, the Market Microstructure Confronting Many Viewpoints #5 Conference, the Future of Financial Information Conference, the Investment Management Research Conference, the Asian Finance Association Annual Meeting, the Financial Research Network (FIRN) Annual Meeting, the FIRN Asset Pricing Workshop, the Australasian Finance and Banking Conference, the J.P. Morgan Quantference, Brigham Young University, Monash University, the University of Western Australia, the University of Sydney, Curtin University, the University of Technology Sydney, and the Stockholm School of Economics in Riga for comments that have helped to improve this paper.

The issue of what drives stock price movements is a fundamental question in finance with implications for understanding risk, informational efficiency, and asset pricing. By understanding the stock return generating process, researchers can address questions such as whether private information is more important than public information, whether the role of market-wide information is increasing or decreasing over time, or how much noise is in stock price movements. A methodology for measuring the return generating process is also useful for evaluating the impacts of recent phenomena such as the growth in passive investing and algorithmic trading, among others. This paper develops a new tool that allows stock returns to be decomposed into various information components while simultaneously allowing for price changes to occur due to non-informational reasons.

There are currently two dominant approaches to decomposing the drivers of stock price changes. One approach exploits the canonical discounted cash flow valuation model to divide a return series into cash flow and discount rate related return components (Campbell and Shiller, 1988a, 1988b; Campbell, 1991). The second decomposes returns into market-wide news and firm-specific news using the R^2 from a regression of stock returns on market returns (Morck, Yeung, and Yu, 2000).

The existing decompositions have limitations, which we overcome with the variance decomposition developed in this paper. The first limitation of existing decompositions arises from ignoring the role of noise in stock returns. Noise can distort the existing measures. For example, an improvement in market efficiency from a reduction in idiosyncratic noise in prices will increase the R^2 measure, yet an increase in the R^2 is usually interpreted as a *decrease* in market efficiency. The second limitation is the inability of existing methods to disentangle information into more refined categories.

We propose a new return variance decomposition model that explicitly accounts for noise and partitions information into various sources. For example, in the baseline model, we decompose the information in stock returns into market-wide information, firm-specific information revealed through trading on private information, and firm-specific information revealed through public news. The framework can be easily extended to further decompose the information components into cash flow and discount rate sub-components or other partitions. Our approach allows for a more nuanced understanding of the specific sources of information that is impounded into stock prices along economically meaningful dimensions.

The empirical model that we propose is based on the permanent-transitory decomposition of Beveridge and Nelson (1981), which has been used in applications ranging from low-frequency macroeconomic business cycles to intraday analysis of financial markets. The intuition for how our decomposition works is as follows. First, in the baseline model we use a vector autoregression (VAR) to measure how a stock's return responds to three shocks including (i) market returns, (ii) firm-specific order flow, and (iii) other firm-specific shocks captured in the stock return residual. We use the VAR impulse response functions to estimate the stock's dynamic price response to each of these shocks. For example, a sudden burst of unexpected buying of a stock typically causes the stock price to temporarily overreact (referred to as "price pressure") and then subsequently revert to a new equilibrium level through time. For each shock, we trace forward the price adjustment process until the price stabilizes. That point provides an estimate of the long-run effect of the shock once the transient effects, including underreaction or overreaction, have dissipated. These long-run impacts, which are the permanent components in the Beveridge and Nelson (1981) decomposition, provide estimates of the information in each shock. For example, the long-run impact of a shock to market returns is the estimated market-wide information in that shock. Similarly, the long-run impact of a shock to firm-specific order flow or a shock to firm-specific returns not associated with order flow (the stock return residual) is the estimated firm-specific private information and the firm-specific public information contained in those shocks, respectively. Conversely, the transient returns from each shock, which cause prices to depart from their efficient or equilibrium levels, make up the estimated noise in prices.

We estimate the model using daily returns on all common stocks listed on the NYSE, AMEX, and NASDAQ between 1960 and 2015, performing the variance decomposition separately for every stock in every year. Overall, we find that roughly 31% of daily return variance is noise. Firm-specific information accounts for the majority (61%) of stock return variance, with market-wide information accounting for the remaining 8% of variance in the full sample. We further partition firm-specific information and find that public firm-specific information plays a larger role than private firm-specific information that is impounded into prices through trading (37% and 24% of variance, respectively). While the estimates suggest that noise makes up an economically meaningful share of daily stock return variance, the estimate is substantially lower than estimates of noise at intra-day horizons (82%).[1]

We find substantial time-series variation in the components of variance. Our results indicate that market efficiency is not static, but rather it is dynamic and the dynamics are heavily influenced by the environment.[2] Some key trends stand out. First, noise increases from the 1970s to the mid-1990s, particularly around a period of collusion by dealers that widened bid-ask spreads. The subsequent decline in noise corresponds to a period of general improvement in liquidity and exogenous decreases in tick sizes (minimum price increments). Separating the sample by firm size shows that the noise component decreases monotonically with firm size. Larger firms have less noise in their prices, as expected. When separating the sample by industry we observe only minor fluctuations in the different components of variance, suggesting that the findings are not specific to a particular industry, nor are they driven by a certain segment of the economy.

Second, the role of firm-specific information has increased through time, driven largely by increases in the amount of public firm-specific information that is reflected in prices. This trend is consistent with increasing informational efficiency through time, which one may expect given a variety of regulations such as the Regulation Fair Disclosure (2000) and the Sarbanes Oxley Act (2002) that have increased the quality and quantity of corporate disclosure.

Third, while market-wide information tends to spike during crises, at other times it is generally not a substantial driver of individual stock returns and typically accounts for around 5–15% of stock return variance. The time-series results shed some light on recent concerns that the rapid growth in passive investing could harm the firm-specific information in stock prices (e.g., Cong and Xu, 2017) and so too

[1] Extrapolating from Hasbrouck (1991b, 1993) for the year 1989.
[2] Because we re-calculate the decomposition each year for each stock, our measure is able to vary along with changes in market structure.

could the growth in high-frequency trading (e.g., Baldauf and Mollner, 2020). Our results indicate that the proportion of firm-specific information in prices has continued to increase throughout the period of growth in passive investing and algorithmic trading.

We validate the results from our methodology by conducting four tests. First, we test hypotheses of how the variance components are expected to relate to various stock characteristics. Second, we examine the effect of exogenous tick size reductions on information and noise. Third, we examine how exogenous shocks to analyst coverage impact public firm-specific information. Finally, we examine mutual fund trading shocks driven by investor redemptions and how they impact noise and information. In all four tests our results comport to the well-documented results in the literature, providing support that the measures behave as we expect. We also compare our information measures to the Hou and Moskowitz (2005) delay measure and the French and Roll (1986) variance ratio and find again that our measures behave as expected.

Being able to accurately measure the amount and type of information in asset prices is vital to understanding the impacts of recent trends and innovations in finance. This paper makes a methodological contribution to the literature by developing a richer and more general variance decomposition that allows the separation of variance into multiple information and noise components. In decomposing information, the key assumptions in our baseline model are that market returns capture market-wide information, allowing for noise, and that the permanent impact of trading in the stock captures private information, also allowing for noise. While these assumptions follow the theoretical and empirical literature[3], they are debatable and researchers can use the framework with alternative assumptions, leading to different information components. Incorporating alternative assumptions is simply a matter of changing the empirical shock proxies in the VAR. For example, public information can cause price changes independent of trading and price changes due to trading on the public information. Whether the latter should be regarded public or private information is a subjective matter—one interpretation (the one taken in this paper's decomposition)

[3] For example, in market microstructure theory models (e.g., Kyle, 1985; Glosten and Milgrom, 1985) the permanent price impact of signed trading volume reflects private information, and this notion is used in empirical models that separate public and private information, such as Hasbrouck (1991a, 1991b).

is that the trading reveals private interpretations of the public information and is therefore counted as private information, yet another interpretation is that because the trading was triggered by public information its impact should be labelled public information. Also, the framework can be easily adapted to other information partitions also by changing the empirical shock measures in the VAR.

Methodologically, our paper is related to Hasbrouck (1993) who pioneers the application of the Beveridge and Nelson (1981) decomposition to intraday financial market data. Hasbrouck uses the decomposition to quantify the noise in trade prices as a summary measure of market quality. Similar to Hasbrouck (1993), we also separate information and noise. Unlike Hasbrouck (1993), we apply our approach to daily returns that are more broadly available and allow a long-horizon examination of changes in the informational characteristics of stock prices. Also in contrast to Hasbrouck (1993), we partition information into different types (market-wide, public, private) and quantify their contributions to variance.

Our method is also related to papers that decompose the information in stock prices. For example, Campbell and Shiller (1988a, 1988b), Chen and Zhao (2009), Chen, Da, and Zhao (2013), and Campbell (1991) decompose return variance into cash flow news and discount rate news. Roll (1988) and Morck, Yeung, and Yu (2000) perform simple decompositions of returns into market-wide variations and firm-specific variations. Similar to these papers, we apply our decomposition to isolate market-wide and firm-specific information. But unlike the existing papers, we separate information from noise and show that this separation is important because noise otherwise biases inference about the types of information in prices. We also partition information into a more granular set of components.

Our study is related to a growing body of literature on the impact of noise in the return process. For instance, Blume and Stambaugh (1983) and Asparouhova et al. (2013) show that noise at daily frequencies causes an economically meaningful bias in returns, equal to 50% or more of the corrected estimate and explains much of the size effect. Jegadeesh (1990) and Lehmann (1990) document significant reversals in stock returns at monthly and weekly horizons, respectively, also consistent with the notion that daily, weekly, and monthly returns contain substantial noise. Asparouhova et al. (2010) find that noisy prices lead to biases in intercept and slope coefficients obtained in any OLS regression using return as the dependent variable. In addition, an extant literature considers noise in prices by examining price efficiency measures such as autocorrelations, variance ratios, reversal strategies, delay measures, post-earnings drift, profitability of momentum strategies, and intraday return predictability based on past order flow or past returns (e.g., Rosch, Subrahmanyam, and van Dijk, 2017; Griffin, Kelly, and Nardari, 2010). Our model provides a convenient method to estimate noise and is a tool that will allow future research to more systematically examine the drivers and effects of noise.

Finally, this paper fits with a growing body of literature studying the price efficiency of markets over time. The results of our work complement those found in Bai, Philippon, and Savov (2016) and Farboodi, Matray, Veldkamp, and Venkateswaran (2021). Bai et al. (2016) show that in more recent times stock prices of S&P 500 firms have become more informative in that they more accurately predict future cash flows. Farboodi et al. (2021) also focus on price informativeness. Both papers show that the improvements in price informativeness have been concentrated in large, growth stocks. Farboodi et al. (2021) go on to introduce a new structural method to measure price informativeness. Our paper takes a different approach, using the variance decomposition technology to disentangle different types of information and noise. Whereas the related papers focus on the ability of stock prices to incorporate low frequency fundamental information, the methodology this paper develops captures higher frequency information from different sources, as well as the noise process. The difference in information and horizons may explain why we find that the noise component of small stocks declines the most, whereas Farboodi et al. (2021) and Bai et al. (2016) show that the tendency for stock prices to reflect quarterly fundamental information has improved the most for large stocks. The difference highlights the role this paper's methodology has in capturing information at higher frequencies and from different sources beyond quarterly and annual financial reports.

1. Empirical model for variance decomposition

This section sets out the empirical model that we use to separate noise and various sources of information. It begins by explaining the motivation for the way we choose to partition information, noting that the framework can be adapted to alternative partitions. It then explains the intuition for how the empirical decomposition separates information and noise, followed by the details of the empirical model.

1.1. Model motivation

There are many sources of information in prices and therefore also many ways that the information can be partitioned. For example, there is market-wide information and stock-specific information, public information and private information, cash flow information and discount rate information, information from company disclosures and information from other sources such as media, overnight information and intraday information, and so on. While the variance decomposition framework that we propose can be adapted to partition information in any of these different ways, we illustrate the approach by choosing a particular partition based on two dimensions of information that have long been of interest in finance and can help understand the evolving informativeness of stock markets.

The first is the distinction between market-wide and firm-specific information. This dimension has long been of interest because firm-specific information is vital for efficient resource allocation across firms (e.g., Wurgler, 2000). The amount of firm-specific information in prices is considered a key measure of the effectiveness of a stock market in its informational role and determines the "real effects" of financial markets (e.g., Bond, Edmans, and Goldstein, 2012). It is also a key determinant of diversification benefits in portfolios of stocks and systematic risk (e.g., Savor and Wilson, 2014).

The second is the distinction between public information such as information disseminated by companies and private information such as proprietary analysis by fundamental investors. In market microstructure theory models, information in prices is either part of the public information set or is private information that gets impounded into prices through the course of trading (e.g., Kyle, 1985; Glosten and Milgrom, 1985). This distinction is important because the amount of private information in prices determines the extent to which corporate managers can use stock prices to learn about their own firms' fundamentals and incorporate this information in the corporate investment decisions (e.g., Chen, Goldstein, and Jiang, 2007). The amount of public information in prices gives a measure of the quality of the disclosure environment, which can help evaluate substantial changes in the regulation of corporate information disclosure in recent decades (e.g., Regulation Fair Disclosure (2000) and the Sarbanes Oxley Act (2002)). Such regulations could result in better disclosure crowding out private information acquisition, or an improvement in the informativeness of prices.

Therefore, we choose to split information into market-wide and stock-specific components and further split firm-specific information into public and private information, although as we noted other partitions are possible and can be accommodated by the variance decomposition approach that we propose. The requirements for selecting an information partition are twofold. First, the shocks to the information set must be measurable because an empirical proxy for all but one of the information sources must be included in the VAR with the omitted information source captured in the residual. Second, the information components must be complete in that they sum to total information.

To show how these sources of information translate into return variances, we derive a modified version of the Jin and Myers (2006) model. In their original model, the intrinsic value of the firm is the present value of future operating cash flows. Cash flows are affected by market-wide and firm-specific information shocks. We extend this setup by injecting noise into the model as one of the shocks to cash flow information.[4] We also split firm-specific information into a part that is revealed through trading on private information, and a part revealed through public information such as company announcements and news.

With these extensions to the Jin and Myers (2006) model, stock return variance is made up of four distinct sources of variation, which can be expressed as four variance shares:

$$\eta_j = \frac{Var(\varepsilon_{j,t})}{Var(\varepsilon_{1,t} + \varepsilon_{2,t} + \varepsilon_{3,t} + \varepsilon_{4,t})}, \tag{1}$$

[4] This could be interpreted as taking the perfect information that is provided to the market in the original model and making it imperfect due to noise such as estimation errors. It could also be interpreted as the addition of noise traders to the model in the spirit of Black's (1986) notion that "noise trading is trading on noise as if it were information." Finally, the ultimate result of injecting noise is that returns vary around the efficient returns, which in real markets can occur due to frictions such as a discrete pricing grid, non-synchronous trading, or imperfect liquidity leading to temporary price pressure.

with $j = \{1, 2, 3, 4\}$, corresponding to variance driven by market-wide information (η_1), private firm-specific information revealed through trading (η_2), public firm-specific information (η_3), and noise (η_4). The extended version of this model and the proofs are in the Internet Appendix Section 8.

For simplicity, in the extension of the Jin and Myers (2006) model above we assume like in the original model each of these components are independent, but in the empirical model we allow information and noise to be correlated. Theoretical (Collin-Dufresne and Fos, 2016) and empirical (Collin-Dufresne and Fos, 2015) studies show that noise and information are positively related. Noise is needed for markets to be liquid (e.g., Kyle, 1985; Glosten and Milgrom, 1985) and to incentivize information acquisition and thereby make prices informative (e.g., Grossman and Stiglitz, 1980; Kyle 1984, 1989; Admati and Plfeiderer, 1988). In the words of Black (1986): "Noise makes financial markets possible, but also makes them imperfect." Going the other way, an exogenous increase in informed trading that tends to bring more private information into prices at a faster rate, imposes adverse selection risk on liquidity providers, causing a deterioration in liquidity and potentially more noise in prices (e.g., Kyle, 1985; Glosten and Milgrom, 1985). Therefore, the relation between information and noise in theory is symbiotic.

Our empirical results support this positive, bi-directional relation between information and noise. In the main analysis, we report the variance components as shares that are normalized to sum to 100%. As such we are discussing the relative levels of the four components. In the Internet Appendix Section 3, we repeat the analysis in levels. As predicted by Collin-Dufresne and Fos (2016) we show there is a positive relation between the level of noise in returns and the amount of information in returns.

1.2. Intuition for how the empirical decomposition of variance is performed

Stock prices are driven by several sources of shocks—some sources are observable (e.g., trades, market movements) and some are unobservable, other than the effects they have on stock prices. Each shock can convey information and/or contribute noise, including under- or over-reaction to the information in the shocks. Empirically measuring what drives stock price movements involves (i) identifying the desired information partitions (as per the previous section) and corresponding empirical shock measures, (ii) modelling how stock prices respond to the shocks, which we do with a VAR model, and (iii) separating the information conveyed by each shock from the noise created by the shock, which we do with Beveridge Nelson (1981) permanent-temporary decompositions.

We work through an example to illustrate the intuition for how we use the Beveridge Nelson (1981) permanent-temporary decomposition approach to separate information and noise. Consider a simplified,[5] two-equation, one-lag VAR model in which a stock's returns (r_t) are driven by two types of shocks: (i) shocks coming from trading measured by x_t, which is buy-volume minus sell-volume ($\varepsilon_{x,t}$ is the unexpected change in x_t), and (ii) shocks from all other sources, $\varepsilon_{r,t}$:

$$\begin{aligned} r_t &= 0.5r_{t-1} + 0.9x_t - 0.7x_{t-1} + \varepsilon_{r,t} \\ x_t &= 0.1x_{t-1} + \varepsilon_{x,t} \end{aligned} \tag{2}$$

While the coefficients are illustrative, they are chosen so that the VAR has properties seen in our sample.

The standard impulse response functions for this VAR model are shown in Figure 1. Panel A shows the cumulative return to an unexpected unit shock to trading ($\varepsilon_{x,0} = 1$).[6] Prices initially overshoot and gradually revert to a stable level. The overshoot and subsequent reversion reflect a temporary component of returns (noise), whereas the price stabilizing above its initial level reflects a permanent component of the returns (information). The permanent component is the point at which the projected price stabilizes. In the figure, the price stabilizes by $t = 10$, revealing that the permanent component is 0.45, which we denote as θ_x.[7] In this example the units are arbitrary, but could be interpreted as percentage changes in the stock price.

Insert Figure 1 About Here

[5] We have reduced the number of lags, omitted some variables, and assumed some zero coefficients to simplify the illustration, but the mechanics and intuition are the same for more complicated VAR models.
[6] The impulse response functions are calculated in the standard way: assume the system is in a steady state meaning that all past shocks are zero $\varepsilon_{x,t} = \varepsilon_{x,t} = 0 \; \forall t < 0$, then impose the relevant shock at time $t = 0$ (i.e., either $\varepsilon_{x,0} = 1$, $\varepsilon_{r,0} = 1$, or $\varepsilon_{x,0} = \varepsilon_{r,0} = 1$), and finally calculate the estimated values of r_t and x_t going forward from the shock using the estimated model coefficients.
[7] Formally, θ_x is the expected cumulative return (price) response to the unit shock as $t \to \infty$, but in practice this is approximated by the response at a finite t that is sufficiently far forward from the shock that the response has stabilized (e.g., Beveridge and Nelson, 1981).

The economic interpretation of the permanent component is that it captures the information in the shock. The economic interpretation of the transient component is that it captures temporary deviations of prices from their long-run equilibrium levels, referred to as "noise." Such deviations can arise from a variety of sources, such as overreaction to information, underreaction to information, imperfect liquidity, price impacts of noise trading, and microstructure frictions.

The plots on the right-hand side of Figure 1 show how this approach decomposes returns into information and noise. Panel A shows how an overreaction to a buying shock at $t = 0$ generates noise. The dashed line is the information content of the shock—a cumulative return of $\theta_x = 0.45\%$, which is the long-run permanent response to the shock based on the impulse response function on the left. The dots show the price response to the shock.[8] At $t = 0$, the return is $r_0 = 0.90\%$, the estimated information component of the shock is $\theta_x \varepsilon_{x,0} = 0.45\%$, and therefore the transient component of r_0 is 0.45% ($r_0 - \theta_x \varepsilon_{x,0}$). At $t = 1$, if there are no further shocks ($\varepsilon_{x,1} = \varepsilon_{r,1} = 0$) the price begins to correct so we get a return of $r_1 = -0.16\%$. As there is no new information the noise return at $t = 1$ is -0.16% ($r_1 - \theta_x \varepsilon_{x,1}$).

Figure 1 Panel B shows how underreaction to a positive news shock at $t = 0$ generates noise. The shock ($\varepsilon_{r,0} = 1$) may come from an after-hours company earnings announcement, for example. The cumulative impulse response function on the left of Panel B illustrates a slow reaction to the information in the shock—there is an immediate return response of $r_0 = 1\%$ and then positive returns in the subsequent periods as the price converges to the long-run impact of $\theta_r = 2.00$. Again, we can use this estimated long-run impact $\theta_r = 2.00$ to decompose actual returns around the shock into information and noise. The right side of Figure 1 Panel B shows that at $t = 0$, the stock return is $r_0 = 1.00\%$, while the estimated information content of the shock is $\theta_r \varepsilon_{r,0} = 2.00\%$, and therefore the transient component is -1.00% ($r_0 - \theta_r \varepsilon_{r,0}$). The initial underreaction creates noise. At $t = 1$, if there are no further shocks ($\varepsilon_{x,1} = \varepsilon_{r,1} =$

[8] For illustration we assume the actual price changes are equal to those predicted by the impulse response function, but the decomposition proceeds in the same manner with any actual price changes.

0) the price begins to correct, we observe a return of $r_1 = 0.50\%$, but no new information, and therefore the noise return at $t = 1$ is 0.50% ($r_1 - \theta_r \varepsilon_{r,1}$).

So far, we have illustrated shocks that arrive one at a time. The model deals with concurrent shocks in much the same way, simply adding their effects. Figure 1 Panel C combines the two previous shocks: there is a positive shock to buying ($\varepsilon_{x,0} = 1$) and a concurrent positive shock from other sources ($\varepsilon_{r,0} = 1$). The impulse response function on the left side of Panel C is the sum of the two individual impulse response functions (Panels A and B). The overreaction to the trading shock offsets some of the underreaction to the shock from other sources, leading to a more efficient price reaction. The information of the combined shock is the sum of the information in the two shocks ($\theta_x + \theta_r = 2.45$). The right side of Panel C shows that the initial return is only $r_0 = 1.90\%$, implying a noise component in the return equal to -0.55% ($r_0 - \theta_x \varepsilon_{x,0} - \theta_r \varepsilon_{r,0}$). While in this example the noise from the two different shocks is offsetting, noise from different shocks can be reinforcing. For example, suppose the shock to trading was negative (selling) while the shock from other sources was positive, the two noise returns would both be negative.[9]

While the variance decomposition that we apply is a more general and extended version of the examples above, the intuition and mechanics are the same. In the model we apply, the information driven return and noise return at each time are inferred from the actual estimated shocks (the residuals of the VAR model) rather than the illustrative hypothetical one-unit shocks. Also, in the data there are new shocks in almost every period, rather than each shock being followed by a period of calm. Thus, the efficient price, which is the cumulation of the information components, evolves dynamically through time. The stock price fluctuates around the efficient price due to the many transient shocks that temporarily push prices away from their efficient levels.

Figure 1 Panel D illustrates this more realistic scenario of many different shocks arriving sequentially, leading to a dynamic efficient price and an actual price that is continually approaching the

[9] For example, if we have $\varepsilon_{x,0} = -1$, $\varepsilon_{r,0} = 1$, and the actual return is as estimated $r_0 = 0.10\%$, the information component of the return is $\theta_x \varepsilon_{x,0} + \theta_r \varepsilon_{r,0} = 0.45(-1) + 2(1) = 1.55\%$ and the noise component in the return is $-1.45\% = 0.10\% - 1.55\%$, which is much larger in magnitude than in the example in Panel C (-0.55).

efficient price. Despite the multitude of shocks, the decomposition of information and noise at each point in time is the same: the shocks are identified from the VAR, the information content of those shocks is estimated as the long-run permanent effect of those shocks (the magnitude of the shock multiplied by the corresponding θ), giving an estimated efficient price that is the cumulation of the information-driven returns. The pricing errors are the differences between the observed prices and the estimated efficient prices. The noise component of each period's return is the change in the pricing error from one period to the next, which is equivalent to the observed return minus the estimated information-driven return.

1.3. Baseline variance decomposition model

Consider the log of the observed price at time t, p_t, as the sum of two components:

$$p_t = m_t + s_t \ , \tag{3}$$

where m_t is the efficient price and s_t is the pricing error. The pricing errors can have a temporary (short-run) effect on the price, but they do not affect price in the long run (no permanent effect). m_t follows a random walk with drift μ, and innovations w_t:

$$m_t = m_{t-1} + \mu + w_t \ . \tag{4}$$

The innovations reflect new information about the stock's fundamentals and are thus unpredictable, $\mathrm{E}_{t-1}[w_t] = 0$. The drift is the discount rate on the stock over the next period (day). The stock return is:

$$r_t = p_t - p_{t-1} = \mu + w_t + \Delta s_t \ . \tag{5}$$

We partition the information impounded into stock prices into three sources: market-wide information, private firm-specific information incorporated through trading, and public firm-specific information such as firm-specific news disseminated in company announcements and by the media. This partitioning is *complete* in that all information must fall into one of these mutually exclusive categories: information is either market-wide or it is firm-specific, firm-specific information is either publicly available or it is not in which case it is private. But this partitioning is not *unique* in that there are other ways of partitioning information.

The random-walk innovations, w_t, can then be decomposed into three parts:

$$w_t = \theta_{r_m}\varepsilon_{r_m,t} + \theta_x\varepsilon_{x,t} + \theta_r\varepsilon_{r,t} \; , \tag{6}$$

and thus

$$r_t = \underbrace{\mu}_{\text{discount rate}} + \underbrace{\theta_{r_m}\varepsilon_{r_m,t}}_{\text{market-wide info}} + \underbrace{\theta_x\varepsilon_{x,t}}_{\text{private info}} + \underbrace{\theta_r\varepsilon_{r,t}}_{\text{public info}} + \underbrace{\Delta s_t}_{\text{noise}} \; , \tag{7}$$

where $\varepsilon_{r_m,t}$ is the unexpected innovation in the market return and $\theta_{r_m}\varepsilon_{r_m,t}$ is the market-wide information incorporated into stock prices, $\varepsilon_{x,t}$ is an unexpected innovation in signed dollar volume and $\theta_x\varepsilon_{x,t}$ is the firm-specific information revealed through trading on private information, and $\theta_r\varepsilon_{r,t}$ is the remaining part of firm-specific information that is not captured by trading on private information ($\varepsilon_{r,t}$ is the innovation in the stock price). Changes in the pricing error, Δs_t, can be correlated with the innovations in the efficient price, w_t. The identifying assumption that our model inherits from Beveridge and Nelson (1981) is that both the permanent (information) and transient (noise) components are driven by the same shocks. This assumption explicitly allows for correlation between information and noise, consistent with theory.

We estimate the components of Equation (7) using a structural VAR with five lags to allow a full week of serial correlation and lagged effects:

$$\begin{aligned} r_{m,t} &= \textstyle\sum_{l=1}^{5} a_{1,l} r_{m,t-l} + \sum_{l=1}^{5} a_{2,l} x_{t-l} + \sum_{l=1}^{5} a_{3,l} r_{t-l} + \varepsilon_{r_m,t} \\ x_t &= \textstyle\sum_{l=0}^{5} b_{1,l} r_{m,t-l} + \sum_{l=1}^{5} b_{2,l} x_{t-l} + \sum_{l=1}^{5} b_{3,l} r_{t-l} + \varepsilon_{x,t} \\ r_t &= \textstyle\sum_{l=0}^{5} c_{1,l} r_{m,t-l} + \sum_{l=0}^{5} c_{2,l} x_{t-l} + \sum_{l=1}^{5} c_{3,l} r_{t-l} + \varepsilon_{r,t}, \end{aligned} \tag{8}$$

where $r_{m,t}$ is the market return, x_t is the signed dollar volume of trading in the given stock (positive values for net buying and negative values for net selling), and r_t is the stock return.

We use the VAR model to identify the information components as follows. The permanent returns in response to market return shocks provide the estimated market-wide information. The permanent returns in response to signed dollar volume shocks provide the estimated firm-specific private information. And the permanent returns in response to stock return shocks controlling for market returns and signed dollar volume provide the estimated firm-specific public information. Noise is the net transitory return from these three sources of shocks or, equivalently, the return that is left after subtracting the information components. Therefore, the approach effectively starts with a stock return and first identifies market-wide information, leaving behind firm-specific information and noise. It then identifies the private firm-specific information associated with trading to leave behind public firm-specific information and noise. Finally, it identifies the public firm-specific information as the permanent returns that are not associated with market returns or trading, leaving behind the noise component of returns.

Consequently, misspecification in the measure of market movements can cause misestimation of market-wide versus firm-specific information.[10] Similarly, misspecification in the measure of trading can cause misestimation of public versus private firm-specific information. We choose to label information impounded into prices through trading as private information, consistent with empirical market microstructure models (e.g., Hasbrouck, 1991a, 1991b). However, in reality the distinction between public and private information can at times be blurred. For example, public information (e.g., a company announcement) can lead to trading due to private interpretations of the information—while public information was the trigger, the private interpretations also contribute to the price informativeness. Also, some changes to prices quoted by market makers may contain private information but without necessarily triggering trades.

We use the product of price, volume, and the sign of the stock's daily return as a proxy for the signed dollar volume, x_t, similar to Pastor and Stambaugh (2003). We use this proxy because it has minimal data requirements making it applicable in a long time series and across many markets. It is also similar to "bulk volume classification" (see Easley, Lopez de Prado, and O'Hara, 2016), potentially making it more robust to capturing private information in different market types because informed traders can use limit orders (e.g., Kaniel and Liu, 2006). We test an alternative measure of order imbalance using the Lee and Ready

[10] There could be noise in market returns, although we expect less noise at the market level than in individual stock returns. For example, market returns may gradually reflect market-wide information due to delayed reactions of some stocks in the market index. If individual stocks react efficiently to market returns, they will factor in the market's tendency to underreact and their responses could efficiently incorporate market-wide information despite noise in market returns. But if a stock does not efficiently react to the information in market returns, the decomposition will capture the stock's underreaction or overreaction to market returns as noise.

(1991) algorithm to identify the trade initiator and then sum the buyer-initiated trade volume minus the seller-initiated trade volume. This alternative measure produces lower estimates of the amount of private firm-specific information in prices and higher estimates of the amount of public firm-specific information in prices, which we interpret as underestimation of private information.[11]

The lags of stock returns in the VAR account for short-term momentum as well as reversals that can be driven by temporary price impacts from trading (e.g., Hendershott and Menkveld, 2014). The lags of signed dollar volume account for persistence in order flow (e.g., Hasbrouck, 1988). Finally, the lags of market returns account for first-order serial correlation in market returns due to non-synchronous trading (e.g., Scholes and Williams, 1977) as well as delayed stock price reactions to market-wide information (e.g., Hou and Moskowitz, 2005). We choose five lags because for most stocks in most years (more than 75% of the sample), the optimal number of lags according to the corrected Akaike Information Criterion (AIC) is five or less (see Section 4 of the Internet Appendix). There are, however, stocks that require a larger number of lags to fully capture the dynamics, consistent with the findings of Boguth, Carlson, Fisher, and Simutin (2016). We therefore re-estimate the variance decomposition using twice as many lags where it is suggested by the AIC and find that our main results are robust to this alternative specification (see Section 4 of the Internet Appendix).

The structural VAR above embeds contemporaneous relations between the variables. First, market-wide information can be reflected in stocks contemporaneously, but because each stock is a small part of the market index, individual stock returns and trades have a negligible contemporaneous impact on the market return. Second, trading activity in a stock can be contemporaneously caused by market returns and can contemporaneously cause changes in the stock price, but not vice versa. To the extent that returns can trigger trading activity contemporaneously (within the same day) or exogenous events such as company announcements can trigger both (trading and returns), the model will tend to overstate the extent to which trading activity drives returns. Therefore, in splitting firm-specific information into public and private

[11] See Section 5 of the Internet Appendix for more details and results from the alternative measure.

components, our structural assumptions tend to estimate the upper bound on private information and lower bound on public information. As a result of explicitly modelling the contemporaneous relations between variables, the structural VAR innovations $\{\varepsilon_{r_m,t}, \varepsilon_{x,t}, \varepsilon_{r,t}\}$ are contemporaneously uncorrelated. Appendix A has the details of how we estimate this structural VAR.

The structural VAR innovations $\{\varepsilon_{r_m,t}, \varepsilon_{x,t}, \varepsilon_{r,t}\}$ are the shocks that contain information and noise. The information in each shock is the estimated long-run effect on the price, which is the permanent component in the Beveridge and Nelson (1981) decomposition. Therefore, the market-wide information at time t is $\theta_{r_m}\varepsilon_{r_m,t}$, where $\varepsilon_{r_m,t}$ is the innovation in the structural VAR (Equation (8)) and θ_{r_m} is the long-run effect of a unit shock, inferred from the cumulative impulse response function of returns to a unit shock $\varepsilon_{r_m,t} = 1$.[12] Similarly, the private firm-specific information revealed through trading and the public firm-specific information revealed through other sources are estimated as $\theta_x\varepsilon_{x,t}$ and $\theta_r\varepsilon_{r,t}$, where $\varepsilon_{x,t}$ and $\varepsilon_{r,t}$ are the innovations in the structural VAR (Equation (8)) and θ_x and θ_r are the long-run cumulative impulse responses of returns to a unit shock $\varepsilon_{x,t} = 1$ and to a unit shock $\varepsilon_{r,t} = 1$.

The sum of the estimated information components gives the innovation in the efficient price, $w_t = \theta_{r_m}\varepsilon_{r_m,t} + \theta_x\varepsilon_{x,t} + \theta_r\varepsilon_{r,t}$. The change in the pricing error (noise return), is the realized return that is not attributable to information or discount rate (drift): $\Delta s_t = r_t - \mu - w_t = r_t - a_0 - \theta_{r_m}\varepsilon_{r_m,t} - \theta_x\varepsilon_{x,t} - \theta_r\varepsilon_{r,t}$, and the noise variance, σ_s^2, is estimated as the variance of the time-series of Δs_t.[13]

Taking the variance of the innovations in the efficient price we get $\sigma_w^2 = \theta_{r_m}^2\sigma_{\varepsilon_{r_m}}^2 + \theta_x^2\sigma_{\varepsilon_x}^2 + \theta_r^2\sigma_{\varepsilon_r}^2$. Recall, the structural model errors are contemporaneously uncorrelated by construction and therefore the covariance terms are all zero. The contribution to the efficient price variation from each of the information

[12] We estimate the long-run (permanent) effect as the cumulative return response at $t = 15$, which is three times the number of lags in the VAR and allows enough time for the return to stabilize in response to a shock.

[13] Recall that the drift in the efficient price (μ) is a constant in this baseline model. It could be estimated by adding constants to the VAR and taking the estimated constant in the return equation (a_0). However, it is not required in the calculation of the noise variance because irrespective of whether it is included in calculating Δs_t, being a constant, it does not affect the variance, σ_s^2.

components is $\theta_{r_m}^2\sigma_{\varepsilon_{r_m}}^2$ (market-wide information), $\theta_x^2\sigma_{\varepsilon_x}^2$ (private firm-specific information), and $\theta_r^2\sigma_{\varepsilon_r}^2$ (public firm-specific information). The estimated components of variance are therefore:

$$
\begin{aligned}
MktInfo &= \theta_{r_m}^2\sigma_{\varepsilon_{r_m}}^2 \\
PrivateInfo &= \theta_x^2\sigma_{\varepsilon_x}^2 \\
PublicInfo &= \theta_r^2\sigma_{\varepsilon_r}^2 \\
Noise &= \sigma_s^2.
\end{aligned}
\tag{9}
$$

Normalizing these variance components to sum to 100% gives variance shares:

$$
\begin{aligned}
MktInfoShare &= \theta_{r_m}^2\sigma_{\varepsilon_{r_m}}^2/(\sigma_w^2+\sigma_s^2) \\
PrivateInfoShare &= \theta_x^2\sigma_{\varepsilon_x}^2/(\sigma_w^2+\sigma_s^2) \\
PublicInfoShare &= \theta_r^2\sigma_{\varepsilon_r}^2/(\sigma_w^2+\sigma_s^2) \\
NoiseShare &= \sigma_s^2/(\sigma_w^2+\sigma_s^2).
\end{aligned}
\tag{10}
$$

Accordingly, $MktInfo$, $PrivateInfo$, $PublicInfo$, and $Noise$ are the variance contributions of market-wide information, trading on private firm-specific information, firm-specific information other than that revealed through trading, and noise, respectively. $PrivateInfoShare$, $PublicInfoShare$, and $MktInfoShare$ are corresponding shares of variance from those various sources of stock price movements.[14] Meanwhile, $NoiseShare$ reflects the relative importance of pricing errors due to overreaction/underreaction to information, illiquidity, price pressures, or other microstructure frictions. This baseline variance decomposition is summarized in Figure 2.

Estimating the variance components and shares given in (9) and (10) conveniently only requires two sets of inputs: (i) the variances of the structural VAR innovations (shocks), $\sigma_{\varepsilon_{r_m}}^2$, $\sigma_{\varepsilon_x}^2$, $\sigma_{\varepsilon_r}^2$ and (ii) the long-run cumulative return responses to these shocks θ_{r_m}, θ_x, and θ_r. We obtain both sets of these inputs from the VAR using the procedure in Appendix A.

[14] We use the sum of the efficient price innovations variance and the noise variance in the dominator (rather than the total variance) so that the normalized shares sum to 100%. In the Internet Appendix Section 9, we show that ignoring the covariance between the efficient price component and Δs in normalizing the variance shares has a negligible effect on the overall estimates of variance shares.

Insert Figure 2 About Here

2. Variance components through time and in the cross-section

This section applies the variance decomposition model to the data. First, we describe the data (Section 2.1) and report the coefficients of the reduced form VAR (Section 2.2), which is one of the first steps in performing the variance decomposition. Next, we report the estimated variance components in the full sample and compare the noise in daily returns to that of intraday returns (Section 2.3). We then characterize how the variance components change through time (Section 2.4) and in the cross-section (Section 2.5).

2.1. Data

Our sample consists of all common stocks listed on the NYSE, AMEX, and NASDAQ.[15] Most of our analyses focus on the period from 1960 to 2015, but we also report a longer timeseries of the variance components starting in 1926. We use daily data on returns, prices, market capitalizations, volumes, and sectors from the Center for Research in Security Prices (CRSP). We obtain financial balance sheet data and year of incorporation from Compustat, analyst coverage from IBES, and institutional holdings from 13F filings. We remove duplicate stock-day observations and observations with missing return, missing volume, or missing price. We remove stock-years that have fewer than 20 valid daily observations to ensure the VAR (which is estimated separately each stock-year) has a sufficient number of observations and we remove stock-years in which any of the variance components are estimated to be zero. Other than these screens, we do not remove stocks based on other characteristics, but rather, to mitigate the effects of outliers and extremes, we winsorize the estimated variance components in Equation (9) at the 5% and 95% levels each year. Our results are robust to lower levels of winsorization (e.g., 1% and 99% levels; see Internet

[15] Applying the filters shrcd (share code) = 10 or 11 and primexch (primary exchange) = “A”, “N”, or “Q”. We remove stock-years in which the stock changes primary exchange within the year.

Appendix Section 7). After these filters, our main sample has an average of 4,362 stocks per year and a total of 22,025 stocks. Appendix B contains definitions of the key variables.

2.2. VAR coefficients

As each stock may have different dynamics, and the drivers of stock returns may change through time, we perform the variance decomposition on each stock in each year separately. Therefore, we estimate the VAR every stock-year using daily data. Using one-year periods also mitigates concerns about non-stationarity and allows the relations between variables to change through time.

Before proceeding to the variance decomposition, Table 1 reports reduced-form VAR coefficient estimates averaged across the individual VAR models.[16] Below each average, in parentheses, the Table reports the percentage of negative statistically significant (at 5%) coefficients (first number in the parentheses) and the percentage of positive statistically significant (at 5%) coefficients (second number in the parentheses). The coefficients in the market return equation (Panel A) show a tendency for positive first-order serial correlation in market returns, consistent with the effects of non-synchronous trading (Scholes and Williams, 1977) and slow diffusion of market-wide information. In contrast, lags of other variables (trading in individual securities and individual stock returns) do not explain current market returns. The coefficients in the signed dollar volume equation (Panel B) show a tendency for buying to follow positive market returns. They also show positive serial correlation in daily signed dollar volume, consistent with persistence in order flow (e.g., Hasbrouck (1988) and many subsequent studies).

The coefficients in the individual stock return equation (Panel C) show that stock returns tend to be positively related to lagged market returns consistent with the known slow diffusion of market-wide information (e.g., Hou and Moskowitz, 2005). They also show a tendency in some stock-years for trading (innovations in signed dollar volume) to impact returns not only contemporaneously but also with a lag,

[16] The variance decomposition is based on the structural VAR in Equation (8), but for practical reasons we follow the common practice of first estimating the reduced-form VAR and then using the coefficients and reduced form error covariances to compute the structural VAR impulse response functions and structural VMA. These structural VAR impulse response functions and structural VMA are then used for the variance decomposition.

suggesting that at times the information in trading takes more than one day to be fully reflected in prices. They also indicate the presence of negative serial correlation in stock returns at daily frequencies out to approximately four days, consistent with reversals of pricing errors due to price pressure (e.g., Hendershott and Menkveld, 2014). The VAR coefficient estimates support the use of five lags in the VAR because by the fifth lag very few coefficients are statistically different from zero (besides the 5% that would be expected by chance at the 95% confidence level).

In addition to the average coefficients, which reveal the lead-lag relations between variables, Table 1 also reports the average correlations of the reduced-form VAR residuals for pairs of variables. These correlations reveal the contemporaneous relations between innovations in the variables. Innovations in signed dollar volume are contemporaneously correlated with returns of individual stocks and with market returns. These correlations are consistent with buying pressure pushing prices up as well as positive returns inducing buying (and vice versa for negative returns). There is also a positive contemporaneous correlation between individual stock returns and market returns consistent with individual stocks contributing to the market return but also market returns reflecting market-wide information, which is impounded in individual stock prices.

Insert Table 1 About Here

2.3. Estimates of variance components

Table 2 Panel A reports the estimated variance shares from the baseline model above for the main sample (all US stocks from 1960 to 2015). Recall the variance decomposition is performed separately for every stock-year. From the stock-year estimates we calculate variance-weighted averages of each component.[17] The results show that market-wide information is the smallest component and accounts for around 9% of stock return variance, while firm-specific information accounts for 61% (summing

[17] We use variance-weighted averages in our baseline results for comparability with Morck et al. (2000, 2013). The motivation is that we are trying to understand the drivers of return variance and so want to emphasize stocks that have more variance. In the Internet Appendix (Section 2) we repeat the key tables and figures using equal-weighted means.

$PrivateInfoShare$ and $PublicInfoShare$). Most of the firm-specific information is impounded in prices through public information (37% of variance), while firm-specific private information that is impounded through trading accounts for around 24% of variance. Finally, noise accounts for a substantial 31% of overall daily stock return variance. Using equal-weighted averages, market-wide information is still the smallest component (13%), while firm-specific information accounts for the largest share of variance (63%), and noise is substantial, although slightly smaller (23%). The confidence intervals are formed from the stock-year observations and indicate little uncertainly about the market-wide averages of the variance shares.[18]

Insert Table 2 About Here

Before exploring the time-series and cross-sectional patterns in these variance shares, we consider how these estimates, in particular the noise in returns, compare to other estimates. The first comparison is with intraday returns (e.g., trade-to-trade), which is where similar permanent-transitory decompositions were first used to separate noise from information. Extrapolating from Hasbrouck (1991b, 1993), the implied noise share in intraday trade-to-trade returns is around 82% in US stocks in 1989.[19] For direct comparison, in the year 1989, our model estimates that the noise share in daily returns is around 31%. Therefore, the estimates of the level of noise in daily returns are considerably smaller than estimates of the noise in intraday trade-to-trade returns. One of the reasons for why there is less noise in daily returns than in trade-to-trade returns is that some sources of noise, such as the bid-ask spread, do not scale up when the

[18] In many applications of the variance decomposition, standard errors can be measured in aggregate estimates (e.g., market-wide averages, averages for groups of stocks, or regressions) by treating each stock-year as an individual observation of the variance components, as is done in the analyses in this paper. If, however, one needs a standard error on a single stock-year estimate of the variance components then the suggested approach is to use a bootstrap procedure suitable for time-series models, such as the recursive bootstrap, the moving block bootstrap, or the stationary bootstrap (e.g., Li and Maddala, 1996).

[19] Hasbrouck (1991, 1993) does not report a noise share comparable to ours but we are able to calculate one from his results as follows. The estimated variance of pricing errors in Hasbrouck (1993) is 10.89×10^{-6}, whereas the variance of random walk innovations in Hasbrouck (1991b) using the same sample is 4.7×10^{-6}. If we conservatively assume zero serial correlation in pricing errors (such an assumption will underestimate the noise variance in the following calculation), then the variance of changes in pricing errors is $2\times10.89\times10^{-6}$ (it is the variance of *changes* in pricing errors that adds noise to returns, not just variance of pricing errors). Now if we compute an implied noise share we get 82.17%.

return horizon is increased, yet fundamental volatility (the variance attributable to information) does scale up with the return horizon (this intuition is exploited in the Corwin and Schultz (2012) effective spread estimator). For example, a one-minute return between two successive trades can contain a whole bid-ask spread (if one trade occurs at the bid and the other at the offer) and one minute of fundamental volatility, while a one-day return can also contain a whole bid-ask spread (if one close occurs at the bid and the other at the offer) but a much larger 24 hours of fundamental volatility. Note, however, that bid-ask bounce is only one of several sources of noise in prices.

Another point of comparison is with the noise induced in daily returns by "price pressure," that is, temporary deviations from efficient prices due to risk-averse liquidity providers being unwilling to provide unlimited liquidity. Recently, Hendershott and Menkveld (2014), using data on New York Stock Exchange (NYSE) intermediaries, estimate that at daily frequencies the distortions in midquote prices caused by price pressure (i.e., separate from the effect of bid-ask-bounce) are economically large (0.49% on average) and have a half-life of 0.92 days. The ratio of price pressure (in the midquote) to the variance of the efficient midquote price is 0.33 or 33% in their sample of 697 NYSE stocks during 1994–2005. This ratio of one source of noise to the estimated efficient price volatility is similar in magnitude to the estimated noise share of variance in our model.

Similarly, but at monthly frequencies and using a different approach, Hendershott et al. (2011) estimate that one-quarter of monthly return variance in NYSE stocks is due to transitory price changes that are themselves partially explained by cumulative order imbalances and market-makers' inventories (price pressure). Again, this is just one source of noise and in monthly returns, but it is also close in magnitude to our estimate.

Finally, our finding that a considerable proportion of the variance in daily returns is noise is consistent with studies such as Jegadeesh (1990) and Lehmann (1990) who document significant predictability (reversals) in stock returns at one month and one-week horizons, respectively. Avramov, Chordia, and Goyal (2006) and Nagel (2012) show that the reversals reflect deviations from efficient prices. They find that non-informational demand generates price pressure that is reversed once liquidity suppliers react to potential profit opportunities and the uninformed demand for liquidity abates. While it is difficult to express the reversals documented by these studies as a percentage of variance to directly compare them to our estimates of noise, Jegadeesh (1990), Lehmann (1990), Avramov et al. (2006), and Nagel (2012) show that the price distortions involved in reversals are economically meaningful, consistent with the economically meaningful noise share estimated by our model. Similarly, Asparouhova et al. (2013) show that noise at daily frequencies causes an economically meaningful bias in returns, equal to 50% or more of the corrected estimate.

2.4. Variance components through time

Figure 3 shows how the stock return variance components change through time. Panel A shows the extended sample period from 1926 to 2020 and Panel B zooms in on the period from 1960 to 2015, which is used in subsequent analyses that are limited by the availability of other datasets. There are several noteworthy long-term trends. First, the share of noise in prices has declined from around 40% of variance in the 1920s and 1930s to 30% of variance in the 1960s and down to around 20% of variance recently, although not monotonically. Noise rose through the 1990s, spiking in 1997, and has gradually declined since then. The high levels of noise in prices in the 1990s are at least partly driven by collusive behavior of dealers during that period, which involved effectively widening the tick size by avoiding odd-eighth quotes and thereby increasing bid-ask bounce (Christie and Schultz, 1994).

Table 2 Panel B confirms that stock returns after 1997 tend to have a smaller proportion of noise and higher information content. The differences between the two sub-periods are statistically significant as well as economically meaningful. For example, the average noise share decreases from 35.42% in the period 1960 to 1997 to only 25.78% after 1997.[20] In the Internet Appendix (Table A.3) we show that the *level* of noise is also lower after 1997 (although the decrease in the *level* is not statistically significant) and the *level*

[20] The Internet Appendix (Section 6) reports results from difference-in-differences models that show that the dealer collusion substantially increased the amount and share of noise in prices. We provide more detailed analysis of the effects of tick size reductions in Section 3.1.

of variance associated with information is higher after 1997 indicating that the decline in the noise share is driven by both (i) a decrease in the level of noise and (ii) an increase in the level of information.

Insert Figure 3 About Here

Second, while noise has declined through time, firm-specific information has become an increasingly important component of stock return variance. Together, the two firm-specific information components have increased from around 50% of variance in the period up to the early 1960s to above 70% of stock return variance in recent years. Table 2 Panel B confirms that this increase in firm-specific information in recent decades is also statistically significant and in the Internet Appendix (Table A.3 and Figure A.4) we show that this result also holds for the *levels* of variance attributable to information, not just the shares. The general trend is consistent with increasing informational efficiency and informativeness of prices through time.

Interestingly, while public and private firm-specific information contribute approximately equally to stock return variance in the early 1960s, these components diverge in more recent decades with publicly available firm-specific information emerging as the dominant component accounting for around 40% of stock return variance in recent years. The shift to public firm-specific information is consistent with the objectives of a variety of regulations such as the Sarbanes Oxley Act (2002) and Regulation Fair Disclosure (2000) to increase both the quality and quantity of public disclosure by companies.

As a short aside, our estimates of the proportion of firm-specific information that is impounded in prices through trading (private information) compared to public information are similar to Hasbrouck's estimates of the role of trading in impounding new information. Using intraday data, Hasbrouck (1991b) estimates that (in 1989) 34.3% of the information in prices gets impounded via individual trades. Despite differences in model, sample, and frequency, using daily data we estimate the fraction in 1989 to be around 40.8%.[21]

Third, while market-wide information tends to spike during crises, at other times it is generally not a substantial driver of individual stock returns. Throughout the sample period, market-wide information typically accounts for around 5–15% of stock return variance. We find similar trends to those described above when taking equal-weighted averages across stocks (Internet Appendix, Section 2).

The broad trends illustrated in Figure 3 shed light on recent issues concerning the information content of prices. The results show declining noise levels and an increasing dominance of the informational components. These time trends contradict the conclusions from the usual interpretation of the $(1 - R^2)$ measure (e.g., Morck, Yeung, and Yu, 2000), highlighting its shortcomings. For instance, an increase in R^2 is generally interpreted as a decrease in the amount of firm-specific information in prices and therefore reduced market efficiency. Figure 4 shows that the R^2 measure has increased sharply since the mid-1990s, which would imply that market efficiency has been on the decline during the past 20 years. Such a conclusion is at odds with much of the event study literature that suggests different market changes have generally improved market efficiency in recent years (e.g., Brogaard, Hendershott, and Riordan, 2014; Comerton-Forde and Putnins, 2015).

Our variance decomposition explains why the R^2 measure has increased sharply since the mid-1990s: noise in prices (which is largely idiosyncratic) has fallen sharply since the mid-1990s, leading to a rise in the R^2, but *not* a deterioration in market efficiency. The flaw in the R^2 measure is its assumption that all idiosyncratic variation must be firm-specific information. Not only is this assumption false in general, but because the level of noise varies substantially through time and across stocks, inference about market efficiency using the R^2 measure can be severely distorted, as illustrate by Figure 4 (R^2) in comparison to Figure 3 (variance components accounting for noise).

[21] A further difference is that our estimate of 40.8% corresponds to the fraction of *firm-specific* information in prices that is impounded through trading, while the 34.3% corresponds to the fraction of *all* information impounded in prices.

Insert Figure 4 About Here

2.5. Variance components in the cross-section of stocks

We expect that the information and noise in stock returns will vary with stock characteristics. Our first hypothesis is that large stocks will tend to have less noisy prices and therefore their returns will be more informative. Table 2 Panel C reports means of the variance components in size quartiles. Consistent with our hypothesis, large stocks tend to have less noisy prices. Noise declines monotonically with size and the differences across stocks are large.[22] For example, in small stocks, noise accounts for 35.56% of stock return variance, which is about twice that of big stocks at 16.92% of variance. The relatively low level of noise in large stocks is likely driven by a high level of liquidity, making their prices less susceptible to temporary deviations and price pressures.

The results also show that the returns of large stocks are proportionally driven more by information (higher market-wide and private firm-specific information shares). The differences are particularly large for market-wide information, which accounts for 21.09% of the variance in big stocks, but only 5.58% in small stocks. Yet in levels, the total amount of variance attributable to new information is higher in smaller stocks, consistent with the notion that there is more information asymmetry and uncertainty about the valuation of smaller stocks (Internet Appendix Table A.3).

Table 2 Panel D shows that there is considerably less variation across industry groups in what drives stock price movements than across size groups. In variance levels, the HiTech industry stands out as having one of the highest noise variances but also the highest information variance.

Given that the composition of stock return variance differs in the cross-section, particularly by size and to a lesser extent by industry, we examine whether the time-series patterns in variance components are due to market composition changes. The mix of industries in the market has changed through time and

[22] The same result holds in variance levels (Internet Appendix Table A.3).

listed stocks have tended to become larger through time. Therefore, in Figures 5 and 6 we repeat the exercise of plotting the time-series of variance components, but this time by size group and by industry group. We form the size groups with respect to thresholds ($100 million and $1 billion in 2010 dollars) that are inflation adjusted through time, rather than size quartiles to keep the size groups relatively comparable through time even as the market composition changes. Time-series trends in the variance components within size or industry groups are less susceptible to compositional changes than the pooled time-series.

Figure 5 shows that all size groups have a similar trend with respect to market-wide information, including the peaks during crises. Large stock returns consistently reflect a higher share of market-wide information through time than smaller stocks. All size groups show quite similar trends in private firm-specific information, except for a period of temporary divergence in the 1990s. The increase in public firm-specific information through time is driven mainly by smaller stocks consistent with improvements in their corporate disclosure. Finally, noise is consistently higher for smaller stocks and smaller stocks are largely responsible for the decline in noise through time, particularly since the mid-1990s. These general trends also hold for variance levels (Internet Appendix Figure A.4), which show stable levels of market-wide information variance through time, increasing levels of public and private firm-specific information, and a decline in noise since peaking in the late 1990s, although with spikes during crises (such as 2008–2009) and episodes such as the dot com period (1999–2001).

Insert Figure 5 About Here

Figure 6 shows that the variance components in different industry groups display remarkably similar time-series trends. Not only are the long-run trends in the types of information and noise similar across the industry groups, but so too are many of the year-to-year fluctuations. This result indicates that the time-series trends are not driven by changing industry composition in the market. Furthermore, it indicates that much of the variation in the information and noise shares is systematic and not just an artifact of estimation error or random fluctuations. Recall that the variance decomposition is performed separately (independently) for each stock in each year. The commonality in the variance component trends across groups of stocks (in this case industry groups) points to systematic drivers of the type of information and degree of noise in prices.

Insert Figure 6 About Here

3. Validation tests

The preceding section provides some validation of the empirical variance decomposition, showing that the variance components have reasonable time-series and cross-sectional properties. This section tests hypotheses about how the variance components are expected to relate to several stock characteristics and validates the components by comparing them to estimates from other approaches (Section 3.1). We then present further validation tests exploiting exogenous variation in three settings: (i) exogenous shocks to tick sizes (Section 3.2), (ii) exogenous shocks to analyst coverage (Section 3.3), and (iii) exogenous shocks to mutual fund trading (Section 3.4). We also relate the variance components to other measures of noise and information (Section 3.5).

3.1. Variance components and stocks characteristics

We expect that the information and noise in stock returns will vary with stock characteristics. As a validation test, we develop hypotheses for a range of stock characteristics and test their relations to the variance components using panel regressions:

$$Share_{i,t} = \alpha + \gamma_1 D_t^{POST} + \sum_j \gamma_j D_{j,i}^{INDUSTRY} + \sum_k \gamma_k X_{k,i,t} + \sum_m \gamma_m Z_{m,i,t} + \varepsilon_{i,t} \ , \qquad (11)$$

where $Share_{i,t}$ is one of the variance component shares ($MktInfoShare_{i,t}$, $PrivateInfoShare_{i,t}$, $PublicInfoShare_{i,t}$, $NoiseShare_{i,t}$) for stock i in year t. D_t^{POST} is an indicator variable that takes the value of one after 1997 and zero before. $D_{j,i}^{INDUSTRY}$ is a set of industry indicator variables ($D_i^{Consumer}$, $D_i^{Healthcare}$, D_i^{HiTech}, and $D_i^{Manufact}$). $X_{k,i,t}$ is a set of stock characteristics. In the baseline panel regressions, $X_{k,i,t}$ includes the log stock price ($lnP_{i,t}$) and the stock's log market capitalization ($lnMC_{i,t}$)

and in subsequent models we expand this set. Finally, $Z_{m,i,t}$ is a set of informational efficiency measures from other approaches, including $VarianceRatio_{i,t}$ and $Delay_{i,t}$, which we discuss further below. To compare with the variance shares, we also estimate regression (11) using the $(1 - R^2)$ measure of firm-specific information.

Insert Table 3 About Here

First, given the evidence in prior studies, we expect noise to have decreased through time, controlling for other factors, and we expect that large stocks will have less noisy prices. Table 3 Panel A reports results for the baseline panel regressions. They confirm that stock returns in the 1997–2015 part of the sample tend to contain a significantly lower proportion of noise and more public firm-specific information, controlling for other factors. Therefore, the time-series changes in noise and in firm-specific information are not driven simply by firms becoming larger through time. The returns of large stocks and high-priced stocks are significantly more affected by market-wide information. Large stocks also tend to have less noisy prices and reflect relatively more firm-specific private information, controlling for other factors. Among the five industry groups, stocks in the Healthcare and HiTech sectors tend to have the highest private firm-specific information shares and lowest noise shares.

Next, we expand the panel regressions with an extended set of stock characteristics as determinants of the information and noise in prices. First, we add institutional holdings, measured as the percentage of outstanding shares held by institutional investors. A common view in the literature is that institutional investors tend to be the informed traders in markets responsible for impounding information in prices. This view is supported by existing empirical evidence that finds institutional investors contribute to market efficiency through their trading (e.g., Boehmer and Kelley, 2009; Sias and Starks, 1997). We therefore hypothesize that the level of institutional holdings will be positively related to the amount of firm-specific private information in prices and negatively related to the amount of noise in prices. The results from the extended panel regressions in Table 3 Panel B support these hypotheses, showing that $NoiseShare_{i,t}$ is negatively related with the level of institutional holdings ($InstoHold_{i,t}$) and that $PrivateInfoShare_{i,t}$ is positively related with $InstoHold_{i,t}$.

Second, we add a proxy for hard to value stocks. Kumar (2009) shows that hard to value stocks tend to have higher idiosyncratic volatility, higher share trading turnover, and tend to be younger. We combine these three measures used in Kumar (2009) by taking the first principal component ($HardToValue_{i,t}$). Kumar (2009) shows that individual investors' biases are amplified when stocks are more difficult to value, which in turn attracts a high intensity of informed trading. The informed traders exploit the behavioral biases and in doing so impound information into prices through trading. We therefore hypothesize that hard to value stocks will have a relatively high share of private firm-specific information incorporated through trading. It is unclear ex-ante whether noise will be higher or lower for such stocks as it depends on the relative strength of the increased behavioral biases vs the informed trading. The results in Table 3 Panel B show that consistent with Kumar (2009) and our hypothesis, the higher level of informed trading in hard to value stocks is associated with more private firm-specific information being revealed through trading. Noise tends to be lower in such stocks.

Next, we add two measures of trading costs and illiquidity to capture limits to arbitrage. The first is the stock's average effective bid-ask spread ($BidAskSpread_{i,t}$) and the second is Amihud's (2002) illiquidity measure ($ILLIQ_{i,t}$). Greater trading costs and less liquidity are impediments to arbitrage and discourage informed trade. We therefore hypothesize that the bid-ask spread and illiquidity measure will be negatively related to the amount of private firm-specific information in prices. We also expect these two measures to be positively related to the amount of noise not only because they impede arbitrage but also because they capture the level of microstructure noise from effects such as bid-ask bounce. The results support both hypotheses as trading costs / illiquidity have a significant negative relation with $PrivateInfoShare_{i,t}$ and a significant positive relation with $NoiseShare_{i,t}$.

Finally, we add balance sheet leverage ($Leverage_{i,t}$) to the regressions. Jensen and Meckling (1976) argue that highly leveraged firms incur more monitoring costs, which they try to reduce by disclosing more information to the market. More public firm-specific information can in turn crowd out some private information acquisition. We therefore hypothesize that leverage is positively related to the level of public firm-specific information and negatively related to the level of private firm-specific information. Both effects are supported by the regression results in Table 3 Panel B.

We repeat the panel regressions using the variance components measured in levels. The results (in the Internet Appendix Table A.4) are generally consistent with the overall relations described above.

In contrast, the $(1 - R^2)$ measure that has been used in previous studies as a proxy for firm-specific information or market efficiency shows some unusual relations with stock characteristics in the panel regressions (Table 3). For example, the $(1 - R^2)$ measure suggests that the market is less informationally efficient in the more recent time period, that larger stocks have less informationally efficient prices, and that stocks with more institutional holdings have less efficient prices, controlling for other factors. These results are difficult to reconcile with the existing literature and instead support the notion that this measure captures a mix of firm-specific noise as well as firm-specific information.

3.2. *Exogenous shocks to tick sizes*

Chordia, Roll, and Subrahmanyam (2008) show that decreases in tick sizes increase liquidity, thereby stimulating arbitrage activity and informed trading, leading to "increased incorporation of private information into prices". We therefore use the reduction in tick sizes in the US markets from eighths of a dollar to sixteenths of a dollar on June 24, 1997, as a natural experiment to see whether our variance components detect the increase in the information content of prices.[23]

Theoretically, the clearest prediction is that with the decrease in tick size and increase in liquidity, there will be more informed trading and therefore a higher level of information-related variance in prices. In particular, we expect the level of private firm-specific information in prices to increase following the

[23] There was a further reduction in tick sizes from sixteenths to pennies on January 29, 2001. We focus on the first set of tick size reductions as existing evidence suggests the improvement in informational efficiency was larger around that event (Chordia, Roll, and Subrahmanyam, 2008).

tick size reductions. The predictions for the level of noise are theoretically less clear. On one hand, the reduction in tick sizes could reduce noise by mitigating the microstructure effects of bid-ask bounce, given spreads are narrower following the tick size reductions. However, by decreasing the costs of executing small trades, the tick size reductions could increase the activity of uninformed "noise" traders, thereby driving up the overall level of noise in prices. Therefore, we analyze the effects on private information in prices as a validation test and examine how noise was impacted to understand which of the possible opposing effects dominates.

To eliminate confounding time-series effects, we exploit cross-sectional heterogeneity in the effects of the tick size reductions to effectively create a control group. Given the tick size reductions in dollars are the same for all stocks in the cross-section, lower priced stocks are disproportionately more impacted by the tick size reductions because for them the change in the relative tick size is larger. For example, the tick size reduction from eighths of a dollar to sixteenths is 1.25% of the price of a $5 stock, but only 0.125% of the price of a $50 stock.[24] We therefore expect the strongest effects in low-priced stocks, with high-priced stocks serving as an implicit control group.

To exploit this natural experiment, we take a subsample of one year on both sides of the tick size reduction from eighths of a dollar to sixteenths of a dollar (i.e., we take the years 1996 and 1998), and estimate difference-in-differences models that exploit the cross-sectional heterogeneity in the treatment. The highest priced quartile of stocks had the smallest change in relative tick size and therefore serves as a control group against which to measure the impact of the tick size reduction in other price quartiles:

$$VarianceComponent_{it} = \alpha + \beta D_t^{POST} + \gamma_1 D_t^{POST} Q1_i + \gamma_2 D_t^{POST} Q2_i + \gamma_3 D_t^{POST} Q3_i + \rho_1 Q1_i + \rho_2 Q2_i + \rho_3 Q3_i + \varepsilon_{it}, \quad (12)$$

where D_t^{POST} takes the value of one after the tick size reduction (1998) and zero otherwise. $Q1_i$, $Q2_i$, and $Q3_i$ are indicator variables that indicate the price quartile to which the firm belongs. The highest price quartile, $Q4_i$, is the omitted category. We test three dependent variables corresponding to our hypotheses:

[24] Similarly, the tick-to-price, a measure of the pricing grid coarseness, decreases by a larger amount for lower priced stocks.

the amount of private firm-specific information in prices ($PrivateInfo_{i,t}$), the level of noise in prices ($Noise_{i,t}$), and the share of noise in return variance ($NoiseShare_{i,t}$). We also re-estimate the model in Equation (12) using the log price ($lnP_{i,t}$) instead of the price quartile indicators as a robustness test.

Insert Table 4 About Here

In Table 4, Models 1 and 2 show that, consistent with our hypothesis, the amount of private firm-specific information in prices increases substantially in the stocks most impacted by the reductions in tick sizes. For example, the lowest priced quartile of stocks has an increase in private firm-specific information variance of 0.43 compared to the highest priced quartile of stocks, which is a large effect compared to the unconditional mean of this variance component (3.28). The increases are monotonic across the price quartiles, consistent with our hypothesis.

Models 3 and 4 show that the *level* of noise in prices increases as a result of the tick size reductions, while Models 5 and 6 show that the *share* of noise decreases. Taken together, these results suggest that the increase in noise is smaller than the increase in variance that is due to information leading to a lower proportion of noise in prices. The results are consistent with the notion that the decrease in execution costs as a result of tick size reductions increases informed trading / arbitrage and noise trading, with the effects of the informed traders dominating. Therefore, prices become more informative both in an absolute sense (levels) and a relative (shares) sense.

3.3. Exogenous shocks to analyst coverage

Exogenous shocks to analyst coverage provide another natural experiment that changes the information environment for individual stocks. Given that analysts produce information about individual companies and disseminate this information to a variety of market participants, a reduction in analyst coverage is likely to reduce the amount of public firm-specific information in prices. As the information in prices declines, the relative level of noise is likely to increase. Analyst coverage is expected to have little effect on market-wide information. The effect of analyst coverage on private firm-specific information is ambiguous: analyst-generated information that is made available to only some market participants might be impounded in prices through the course of those participants trading on the information (an increase in private information), but it might also crowd-out private information acquisition (a decrease in private information).

To test the impact of analyst coverage on the information and noise in prices, we use brokerage mergers/closures as a source of exogenous variations in analyst coverage. Broker mergers and closures are plausibly exogenous shocks because the termination of coverage is not driven by the characteristics or behavior of the firm (see Hong and Kacperczyk 2010; Kelly and Ljungqvist 2012; Brogaard et al., 2018). We obtain a list of broker mergers and closures that combines the lists from Hong and Kacperczyk (2010), spanning 1984 to 2005, and Kelly and Ljungqvist (2012), spanning 2000 to 2008. Combining these lists, merging with CRSP and IBES (Institutional Brokers' Estimate System) data, and imposing the requirement that both the acquirer and target brokers must provide overlapping coverage for at least one firm before the broker merger (as per Kelly and Ljungqvist, 2012) we have 41 mergers/closures of brokers that occur during the period 1989–2009. Using the mergers/closures data, we calculate the number of exogenous analyst disappearances per stock-year. These mergers and closures result in exogenous coverage shocks to 4,546 firm-year observations.

Using the exogenous analyst coverage shocks, we estimate the following difference-in-differences model:

$$Share_{i,t} = \gamma_i + \delta_t + \beta_1 CoverageShock_{i,t} + \varepsilon_{i,t}, \quad (13)$$

where $Share_{i,t}$ is one of the variance component shares ($MktInfoShare_{i,t}$, $PrivateInfoShare_{i,t}$, $PublicInfoShare_{i,t}, NoiseShare_{i,t}$) for stock i in year t, γ_i and δ_t are stock and time fixed effects, respectively, and $CoverageShock_{i,t}$ is the number of analyst disappearances due to mergers and closures of brokerage houses during the past two years.[25] We estimate the model above using the period from 1987 to 2011 given that the brokerage mergers and closures occur between 1989 and 2009 and we need to observe a two-year trend before and after the analyst disappearances.

The results in Table 5 show that exogenous decreases in analyst coverage are associated with a decline in public firm-specific information and an increase in the noise share of variance. These results are consistent with the notion that analysts produce firm-specific information that is made public and becomes reflected in prices. It also suggests that the public firm-specific information component of variance from our variance decomposition model can detect this change in the information environment. The coefficient estimates indicate that the exogenous disappearance of each analyst is associated with a decline in public firm-specific information equal to around 1.47% of variance (for comparison, the pooled sample mean of public firm-specific information is around 36.67% of variance).

Shocks to analyst coverage have no significant effect on the amount of market-wide information in prices as expected. Neither do they have a significant impact on the amount of private firm-specific information in prices.

Insert Table 5 About Here

3.4. Exogenous shocks to mutual fund trading

Exogenous shocks to trading of individual stocks by mutual funds in response to investor redemptions provide another natural experiment. The general idea, which has been exploited in many studies (e.g., Edmans, Goldstein, and Jiang, 2012), is that investors' decisions to accumulate or divest units in mutual funds are unlikely to be directly related to the market characteristics of individual stocks, yet the subsequent trading by the mutual fund directly affects individual stocks, including the information and noise in prices.

[25] In contrast to standard difference-in-differences models, here the "treatment" can have different magnitudes depending on how many analysts cease their coverage of a given stock. If the number of analyst disappearances is different in year $t-1$ and year $t-2$, we take the maximum of these two values.

Huang, Ringgenberg, and Zhang (2019) show that when mutual funds receive large exogenous redemptions by investors, they do not scale back their positions uniformly, but rather, they disproportionately sell a subset of low-quality stocks that subsequently underperform. Therefore, mutual fund trading in response to flow shocks has an informational component (the disproportional or "discretionary" selling) and a non-informational component (the pro-rata or "expected" selling). Huang et al. (2019) develop a method to decompose mutual fund trading into these two components[26] and they show that discretionary sales are associated with significant, persistent price changes in the underlying stocks consistent with informed trading, while expected sales have significantly smaller effects that are not statistically different from zero. On this basis, we conjecture that an increase in discretionary trading by mutual funds in response to exogenous flow shocks will be associated with an increase in the private firm-specific information share and a corresponding decrease in the noise share, while a shock to expected trading by mutual funds should have little or no effect on the variance components.

We test this conjecture by estimating the following difference-in-differences model:

$$Share_{i,t} = \gamma_i + \delta_t + \beta_1 DiscretionaryTradeShock_{i,t} + \beta_2 ExpectedTradeShock_{i,t} + \varepsilon_{i,t}, \quad (14)$$

where $Share_{i,t}$ is one of the variance component shares, γ_i and δ_t are stock and time fixed effects, respectively, $DiscretionaryTradeShock_{i,t}$ takes the value of one for stock-quarters that experience strong discretionary buying or selling by mutual funds in response to a redemption shock (defined as the bottom and top deciles of mutual fund discretionary trading), and similarly $ExpectedTradeShock_{i,t}$ takes the value of one for stock-quarters that experience strong non-discretionary fire sales by mutual funds due to fund redemptions (defined as the bottom decile of mutual fund non-discretionary trading).[27]

The results in Table 6 support our conjecture. A shock to discretionary trading by mutual funds is associated with a significant increase in the private firm-specific information share and market information

[26] They decompose the trades of fund managers into (i) expected trading, which is what would occur if the fund manager simply prorated flow shocks across each asset in her portfolio and (ii) discretionary trading, which is the actual minus expected trading. The expected and discretionary trading in each stock is summed across all funds with extreme flows.

[27] We are grateful to Huang, Ringgenberg, and Zhang for sharing the code to calculate these components of mutual fund trading shocks.

share and a corresponding decrease in the noise share. Price informativeness improves in response to more discretionary, informed trading by mutual funds. Also consistent with our conjecture and the findings of Huang et al. (2019), expected trading by mutual funds has little effect on stock prices—the impact on firm-specific information and noise is statistically indistinguishable from zero. Therefore, the evidence from mutual fund trading supports the notion that the variance decomposition detects changes in the informational properties of stock returns, in particular the amount of firm-specific private information and noise.

Insert Table 6 About Here

3.5. Relation between variance components and other measures of information and noise

In the final set of validation tests, we examine the relation between the variance components and two other measures of information in prices. The first is the Hou and Moskowitz (2005) delay metric, defined in Appendix B. The $Delay_{i,t}$ metric measures the incremental contribution of lagged market returns in explaining a stock's current returns and therefore captures inefficiency (delay) in how market-wide information is reflected in stock prices. We therefore expect that higher values of $Delay_{i,t}$ should be associated with less market-wide information in prices (lower $MktInfoShare_{i,t}$) and less efficient, noisier prices (higher $NoiseShare_{i,t}$).

Table 3 reports results from panel regressions that test this hypothesis, controlling for a range of other stock characteristics. The results in both panels show that $MktInfoShare_{i,t}$ has a strong, negative association with $Delay_{i,t}$ consistent with the hypothesis. An increase in $Delay_{i,t}$ from zero (full efficiency) to 0.5 (halfway to the maximum inefficiency) is associated with a reduction of $MktInfoShare_{i,t}$ by 10.62% of variance (a large effect, considering the pooled sample mean of $MktInfoShare_{i,t}$ is 8.62%). The results therefore indicate that $MktInfoShare_{i,t}$ reflects the efficiency with which market-wide information is reflected in prices.

The results also indicate that stocks that are inefficient in reflecting market-wide information (higher values of $Delay_{i,t}$) also tend to have noisier prices (statistically significantly higher $NoiseShare_{i,t}$). These results contrast with the $(1 - R^2)$ measure that has been used in the past as a proxy for informational efficiency. The regressions show that the $(1 - R^2)$ measure is positively related to the delay or inefficiency in reflecting market-wide information, which is opposite to what would be expected if the $(1 - R^2)$ measure captured informational efficiency.

The second existing measure that we test is the variance ratio measure used by French and Roll (1986) and Chordia et al. (2011), among others. This measure (defined in Appendix B) is the variance of returns during trading hours divided by the variance of overnight returns and is therefore intended to capture the amount of information that gets impounded into prices through the course of trading. French and Roll (1986) and Chordia et al. (2011) use the ratio as a measure of informational efficiency. We therefore expect that this measure will be positively associated with private and public firm-specific information and negatively associated with the relative amount of noise in prices.

The panel regression results in Table 3, controlling for a range of other stock characteristics, support these hypotheses. In both panels, the variance ratio measure is positively related to $PrivateInfoShare_{i,t}$ and $PublicInfoShare_{i,t}$ and negatively related to $NoiseShare_{i,t}$. This result supports the interpretation of the intraday/overnight variance ratio as an efficiency measure as well as the ability of our variance decomposition to separate information from noise. The $(1 - R^2)$ measure is also positively associated with the intraday/overnight variance ratio, although with smaller absolute t-statistics than for $PrivateInfoShare_{i,t}$ and $PublicInfoShare_{i,t}$.

4. Conclusion

This study decomposes stock return variance to better understand the roles of different types of information and noise in driving stock price movements. We find that a substantial proportion of return variance, 31%, is noise. Firm-specific information accounts for the majority (61%) of stock return variance, with market-wide information accounting for the remaining 8% of variance. We further partition firm-specific information and find that in the full sample, public firm-specific information plays a larger role than private firm-specific information that is impounded into prices through trading.

There is substantial time-series variation in the components of variance, with some key trends standing out. First, noise increases from the 1970s to the mid-1990s, particularly around a period of collusion by dealers to widen bid-ask spreads, and has substantially declined since then. The decline in noise is partly attributable to narrower tick sizes, which reduces bid-ask bounce, and a general improvement in liquidity and increase in turnover. We show that the recent decrease in noise due to improved liquidity is largely responsible for the increasing R^2 of a market model over the past two decades. An important implication is that a lower R^2 is not necessarily associated with more informationally efficient prices, in contrast to the interpretation of R^2 in prior studies.

Second, the role of firm-specific information has increased through time, driven largely by increases in the amount of public firm-specific information that is reflected in prices. This trend is consistent with increasing informational efficiency through time. The increasing importance of public firm-specific information in stock prices is also consistent with a variety of regulatory reforms such as the Regulation Fair Disclosure (2000) and the Sarbanes Oxley Act (2002) aimed at improving both the quality and quantity of corporate disclosure.

Third, while market-wide information tends to spike during crises, at other times it is generally not a substantial driver of individual stock returns and typically accounts for around 5–15% of stock return variance.

While the broad trends in the components of stock return variance may shed light on recent issues concerning the information content of prices, including the impacts of algorithmic trading and passive investing, this paper’s contribution is largely methodological. The framework for variance decomposition developed in this paper can be applied to analyzing these issues and others, due to (i) its ability to isolate noise from information, which is crucial for correctly characterizing the information in prices, and (ii) the ability to obtain higher frequency estimates of variance components, which is important in analyzing effects that vary through time and understanding recent phenomena that require high resolution estimates of the information/noise components of variance. The framework can be easily extended to even more granular partitions of variance, for example, further splitting market-wide information into public and private components, splitting public firm-specific information into official corporate disclosures vs news/media and other sources, or partitioning any of the information components into cash flow and discount rate related parts.

Appendix A: Estimation of the structural VAR

For the variance decomposition (to construct the variance components and variance shares in Equations (9) and (10)), we need two key inputs from the structural VAR in Equation (8): the variance of the innovations (shocks) in each of the variables, $\sigma^2_{\varepsilon_{r_m}}$, $\sigma^2_{\varepsilon_x}$, $\sigma^2_{\varepsilon_r}$, and the long-run cumulative return responses to these shocks (θ_{r_m}, θ_x, θ_r). We obtain these elements using the common practice of first estimating the reduced-form VAR and using the reduced form error covariances to recover the structural VAR parameters of interest.

We estimate the reduced-form version of the VAR:

$$\begin{aligned} r_{m,t} &= a_0^* + \textstyle\sum_{l=1}^{5} a_{1,l}^* r_{m,t-l} + \sum_{l=1}^{5} a_{2,l}^* x_{t-l} + \sum_{l=1}^{5} a_{3,l}^* r_{t-l} + e_{r_m,t} \\ x_t &= b_0^* + \textstyle\sum_{l=1}^{5} b_{1,l}^* r_{m,t-l} + \sum_{l=1}^{5} b_{2,l}^* x_{t-l} + \sum_{l=1}^{5} b_{3,l}^* r_{t-l} + e_{x,t} \\ r_t &= c_0^* + \textstyle\sum_{l=1}^{5} c_{1,l}^* r_{m,t-l} + \sum_{l=1}^{5} c_{2,l}^* x_{t-l} + \sum_{l=1}^{5} c_{3,l}^* r_{t-l} + e_{r,t} \end{aligned} \tag{A.1}$$

saving the residuals ($e_{r_m,t}$, $e_{x,t}$, and $e_{r,t}$) and the variances/covariances of the residuals. Notice that the only difference in the reduced form VAR compared to the structural VAR is that the contemporaneous relations between the variables are not captured by coefficients ($b_{1,0}$, $c_{1,0}$, and $c_{2,0}$) so they will instead cause contemporaneous correlations in the reduced form residuals. Through successive substitution[28], it is possible to rewrite the reduced form residuals as linear functions of the structural model residuals:

$$\begin{aligned} e_{r_m,t} &= \varepsilon_{r_m,t} \\ e_{x,t} &= \varepsilon_{x,t} + b_{1,0}\varepsilon_{r_m,t} = b_{1,0} e_{r_m,t} + \varepsilon_{x,t} \\ e_{r,t} &= \varepsilon_{r,t} + \left(c_{1,0} + c_{2,0} b_{1,0}\right)\varepsilon_{r_m,t} + c_{2,0}\varepsilon_{x,t} = c_{1,0} e_{r_m,t} + c_{2,0} e_{x,t} + \varepsilon_{r,t}. \end{aligned} \tag{A.2}$$

While the structural model residuals are contemporaneously uncorrelated by construction, Equation (A.2) shows that the reduced form model residuals are contemporaneously correlated, and the contemporaneous correlation can be used to infer the structural model residuals. Specifically, we estimate $b_{1,0}$ by regressing

[28] For example, see Lutkepohl (2005).

the reduced form innovation $e_{x,t}$ on $e_{r_m,t}$ (as per the second equation in (A.2)) and we estimate $c_{1,0}$ and $c_{2,0}$ by regressing the reduced form innovation $e_{r,t}$ on $e_{r_m,t}$ and $e_{x,t}$ (as per the third equation in (A.2)).[29]

From the estimated parameters $b_{1,0}$, $c_{1,0}$, and $c_{2,0}$ and the estimated variances of the reduced form residuals ($\sigma^2_{e_{r_m}}$, $\sigma^2_{e_x}$, and $\sigma^2_{e_r}$), we obtain estimates of the variances of the structural model shocks by taking the variance of (A.2) and rearranging[30]:

$$\begin{aligned} \sigma^2_{\varepsilon_{r_m}} &= \sigma^2_{e_{r_m}} \\ \sigma^2_{\varepsilon_x} &= \sigma^2_{e_x} - b^2_{1,0}\sigma^2_{e_{r_m}} \\ \sigma^2_{\varepsilon_r} &= \sigma^2_{e_r} - \left(c^2_{1,0} + 2c_{1,0}c_{2,0}b_{1,0}\right)\sigma^2_{e_{r_m}} - c^2_{2,0}\sigma^2_{e_x}. \end{aligned} \tag{A.3}$$

We estimate the long-run cumulative impulse response functions of the structural model by computing the equivalent reduced form shocks (using Equation (A.2)) and feeding them through the reduced form model:

(i) A structural shock to market returns $\left[\varepsilon_{r_m,t}, \varepsilon_{x,t}, \varepsilon_{r,t}\right]' = [1,0,0]'$ has a reduced form equivalent $\left[e_{r_m,t}, e_{x,t}, e_{r,t}\right]' = \left[1, b_{1,0}, (c_{1,0} + c_{2,0}b_{1,0})\right]'$.

(ii) A structural shock to trading $\left[\varepsilon_{r_m,t}, \varepsilon_{x,t}, \varepsilon_{r,t}\right]' = [0,1,0]'$ has a reduced form equivalent $\left[e_{r_m,t}, e_{x,t}, e_{r,t}\right]' = \left[0,1, c_{2,0}\right]'$.

(iii) A structural shock to the stock returns $\left[\varepsilon_{r_m,t}, \varepsilon_{x,t}, \varepsilon_{r,t}\right]' = [0,0,1]'$ has a reduced form equivalent $\left[e_{r_m,t}, e_{x,t}, e_{r,t}\right]' = [0,0,1]'$. (A.4)

The cumulative return response to each of these shocks at $t = 15$ (point at which the responses are generally stable[31]) gives estimates of θ_{r_m}, θ_x, and θ_r, respectively.

---

[29] From the estimated parameters $b_{1,0}$, $c_{1,0}$, and $c_{2,0}$ it is possible to recover the structural model innovations using: $\varepsilon_{r_m,t} = e_{r_m,t}$, $\varepsilon_{x,t} = e_{x,t} - b_{1,0}e_{r_m,t}$, and $\varepsilon_{r,t} = e_{r,t} - c_{1,0}e_{r_m,t} - c_{2,0}e_{x,t}$. However, it is not necessary to do so as the variance decomposition can be computed from the long-run impulse response estimates and the variances of the shocks as per the steps in this appendix.

[30] Recall the structural model innovations are uncorrelated by construction.

[31] If the VAR is extended to have a larger number of lags or in a setting in which it takes longer than $t = 15$ for the response to stabilize then simply increase the time at which the response is measured.

The final step is to combine the estimated θ_{r_m}, θ_x, and θ_r with the estimated $\sigma^2_{\varepsilon_{r_m}}$, $\sigma^2_{\varepsilon_x}$, $\sigma^2_{\varepsilon_r}$ from Equation (A.3) to get the variance components and component shares using Equations (8) and (9).

Summary of the five-step procedure:

(i) Estimate the reduced form VAR in Equation (A.1), saving the residuals and variance/covariance matrix of residuals;

(ii) Estimate the parameters $b_{1,0}$, $c_{1,0}$, and $c_{2,0}$ from regressions of the reduced form residuals (second and third equations in (A.2));

(iii) Estimate the variances of the structural innovations using Equation (A.3);

(iv) Estimate the long-run (permanent) cumulative return responses to unit shocks of the structural model innovations, θ_{r_m}, θ_x, and θ_r, using reduced form model impulse response functions with the shocks given in Equations (A.4); and

(v) Combine the estimated variances of the structural innovations from step (iii) with the long-run return responses from step (iv) to get the variance components and variance shares following Equations (9) and (10) in the paper.

Appendix B: Variable definitions

The table below provides descriptions and notation for the components of stock return variance and other efficiency measures. Each variable is estimated for each stock in each year using daily observations. When aggregating across stocks, we take variance-weighted averages (as per Morck, Yeung, and Yu, 2000, 2013).

Variable	Notation	Description
Stock return co-movement	R^2	R^2 is estimated by regressing individual daily stock returns on daily market return.
Noise share	$NoiseShare$	The share of stock return variance attributable to noise.
Market-wide information share	$MktInfoShare$	The share of stock return variance attributable to market-wide information.
Private firm-specific information share	$PrivateInfoShare$	The share of stock return variance attributable to trading on private firm-specific information.
Public firm-specific information share	$PublicInfoShare$	The share of stock return variance attributable to public firm-specific information.
Firm-specific information share	$FirmInfoShare$	The share of stock return variance attributable to firm-specific information (sum of $PrivateInfoShare$ and $PublicInfoShare$).
Delay in impounding market-wide information (Hou and Moskowitz, 2005)	$Delay$	Calculated as $Delay_{i,t} = 1 - \frac{R^2_{constrained}}{R^2_{unconstrained}}$ where $R^2_{constrained}$ is the R^2 in a regression of daily stock returns on contemporaneous market returns and $R^2_{constrained}$ is the R^2 in a regression of daily stock returns on ten lags of market returns.
Variance ratio (French and Roll, 1986; Chordia et al., 2011)	$VarianceRatio$	Variance of returns during trading hours (variance of open-to-close returns, $1/n \sum_1^n r^2_{intraday,i,d}$) divided by the variance of overnight returns (variance of close-to-open returns, $1/n \sum_1^n r^2_{overnight,i,d}$).

References

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

Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Asparouhova, E., Bessembinder, H., Kalcheva, I., 2010. Liquidity biases in asset pricing tests. Journal of Financial Economics 96, 215–237.

Asparouhova, E., Bessembinder, H., Kalcheva, I., 2013. Noisy prices and inference regarding returns. Journal of Finance 68, 665–714.

Avramov, D., Chordia, T., Goyal, A., 2006. Liquidity and autocorrelations in individual stock returns. Journal of Finance 61, 2365–2394.

Bai, J., Philippon, T., Savov, A., 2016. Have financial markets become more informative?. Journal of Financial Economics, 122, 625–654.

Baldauf, M., Mollner, J., 2020. High-frequency trading and market performance. Journal of Finance 75, 1495–1526.

Beveridge, S., Nelson, C., 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. Journal of Monetary Economics 7, 151–174.

Blume, M.E., Stambaugh, R.F., 1983. Biases in computed returns: An application to the size effect. Journal of Financial Economics 12, 387–404.

Black, F., 1986. Noise. Journal of Finance 41, 528–543.

Boehmer, E., Kelley, E.K., 2009. Institutional investors and the informational efficiency of prices. Review of Financial Studies 22, 3563–3594.

Boguth, O., Carlson, M., Fisher, A., Simutin, M., 2016. Horizon effects in average returns: The role of slow information diffusion. Review of Financial Studies 29, 2241–2281.

Bond, P., Edmans, A. and Goldstein, I., 2012. The real effects of financial markets. Annual Review of Financial Economics 4, 339–360.

Brogaard, J., Shi, W., Wei, K., You, H., 2018. Do analysts improve investment efficiency?. Working paper.

Brogaard, J., Hendersott, T., Riordan, R., 2014. High-frequency trading and price discovery. Review of Financial Studies 27, 2267–2306.

Campbell, J.Y., 1991. A variance decomposition for stock returns. Economic Journal 101, 157–179.

Campbell, J.Y., Ammer, J., 1993. What moves the stock and bond markets? A variance decomposition for long-term asset returns. Journal of Finance 48, 3–37.

Campbell, J.Y., Shiller, R.J., 1988a. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228.

Campbell, J.Y., Shiller, R.J., 1988b. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661–676.

Chen, L., Da, Z., Zhao, X., 2013. What drives stock price movements?. Review of Financial Studies 26, 841–876.

Chen, Q., Goldstein, I. and Jiang, W., 2007. Price informativeness and investment sensitivity to stock price. Review of Financial Studies 20, 619–650.

Chen, L., Zhao, X., 2009. Return decomposition. Review of Financial Studies 22, 5213–5249.

Christie, W., Schultz, P., 1994. Why do NASDAQ market makers avoid odd-eighth quotes?. Journal of Finance 49, 1813–1840.

Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and market efficiency. Journal of Financial Economics 87, 249–268.

Chordia, T., Roll, R., Subrahmanyam, A., 2011. Recent trends in trading activity and market quality. Journal of Financial Economics 101, 243–263.

Collin-Dufresne, P., Fos, V., 2015. Do prices reveal the presence of informed trading?. Journal of Finance 70, 1555–1582.

Collin-Dufresne, P., Fos, V., 2016. Insider trading, stochastic liquidity, and equilibrium prices. Econometrica 84, 1441–1475.

Comerton-Forde, C., Putnins, T., 2015. Dark trading and price discovery. Journal of Financial Economics 118, 70–92.

Cong, L.W., Xu, D.X., 2017. Rise of factor investing: Asset prices, informational efficiency, and security design. Working paper.

Corwin, S.A., Schultz, P., 2012. A simple way to estimate bid-ask spreads from daily high and low prices. Journal of Finance 67, 719–759.

Easley, D., Lopez de Prado, M., O'Hara, M., 2016. Discerning information from trade data. Journal of Financial Economics 120, 269–285.

Edmans, A., Goldstein, I., Jiang, W., 2012. The real effects of financial markets: The impact of prices on takeovers. Journal of Finance 67, 933–971.

Farboodi, M., Matray, A., Veldkamp, L., and Venkateswaran, V., 2021. Where has all the data gone?. Review of Financial Studies, forthcoming.

French, K., Roll, R., 1986. Stock return variances: The arrival of information and the reaction of traders. Journal of Financial Economics 17, 5–26.

Glosten, L.R., Milgrom, P.R., 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.

Griffin, J., Kelly, P., Nardari, F., 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. Review of Financial Studies 23, 3225–3277.

Grossman, S.J., Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. American Economic Review 70, 393–408.

Hasbrouck, J., 1988. Trades, quotes, inventories and information. Journal of Financial Economics 22, 229–252.

Hasbrouck, J., 1991a. Measuring the information content of stock trades. Journal of Finance 46, 179–207.

Hasbrouck, J., 1991b. The summary informativeness of stock trades: An econometric analysis. Review of Financial Studies 4, 571–595.

Hasbrouck, J., 1993. Assessing the quality of a security market: A new approach to transaction-cost measurement. Review of Financial Studies 6, 191–212.

Hendershott, T., Li, S., Menkveld, A.J., Seasholes, M.S., 2011. Risk sharing, costly participation, and monthly returns. Working paper.

Hendershott, T., Menkveld, A., 2014. Price pressures. Journal of Financial Economics 114, 405–423.

Ho, T., Stoll, H.R., 1981. Optimal dealer pricing under transaction and return uncertainty. Journal of Financial Economics 9, 47–73.

Hong, H., Kacperczyk. M., 2010. Competition and bias. Quarterly Journal of Economics 125, 1683–1725.

Hou, K., Moskowitz, T.J., 2005. Market frictions, price delay, and the cross-section of expected returns. Review of Financial Studies 18, 981–1020.

Huang, S., Ringgenberg, M.C., Zhang, Z., 2019. The information in asset fire sales. Working paper.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. Journal of Finance 45, 881–898.

Jensen, M.C., Meckling, W.H., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics 3, 305–360.

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

Kaniel, R. and Liu, H., 2006. So what orders do informed traders use?. Journal of Business 79, 1867–1913.

Kelly, B., Ljungqvist, A., 2012. Testing asymmetric-information asset pricing models. Review of Financial Studies 25, 1366–1413.

Kumar, A., 2009. Hard-to-value stocks, behavioral biases, and informed trading. Journal of Financial and Quantitative Analysis 44, 1375–1401.

Kyle, A.S., 1984. Market structure, information, futures markets, and price formation, in International agricultural trade: Advanced readings in price formation, market structure, and price instability, ed. by Gary G. Storey, Andrew Schmitz, and Alexander H. Sarris. Boulder and London: Westview Press, 45–64.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Kyle, A.S., 1989. Informed speculation with imperfect competition. Review of Economic Studies 56, 317–355.

Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–746.

Lehmann, B., 1990. Fads, martingales, and market efficiency. Quarterly Journal of Economics 105, 1–28.

Li, H., Maddala, G.S., 1996. Bootstrapping time series models. Econometric Reviews 15, 115–158.

Lütkepohl, H., 2005. New introduction to multiple time series analysis. Springer Science & Business Media.

Nagel, S., 2012. Evaporating liquidity. Review of Financial Studies 25, 2005–2039.

Morck, R., Yeung, B., Yu, W., 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, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Roll, R., 1988. R^2. Journal of Finance 43, 541–566.

Rosch, D., Subrahmanyam, A., van Dijk, M., 2017. The dynamics of market efficiency. Review of Financial Studies 30, 1151–1187.

Savor, P., Wilson, M., 2014. Asset pricing: A tale of two days. Journal of Financial Economics 113, 171–201.

Sias, R.W., Starks, L.T., 1997. Return autocorrelation and institutional investors. Journal of Financial Economics 46, 103–131.

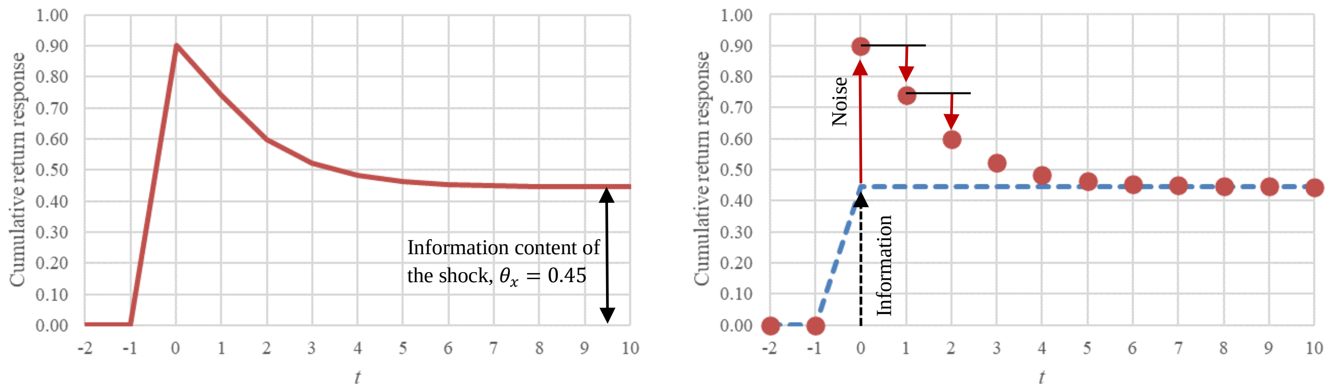
Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data. Journal of Financial Economics 5, 309–327.

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

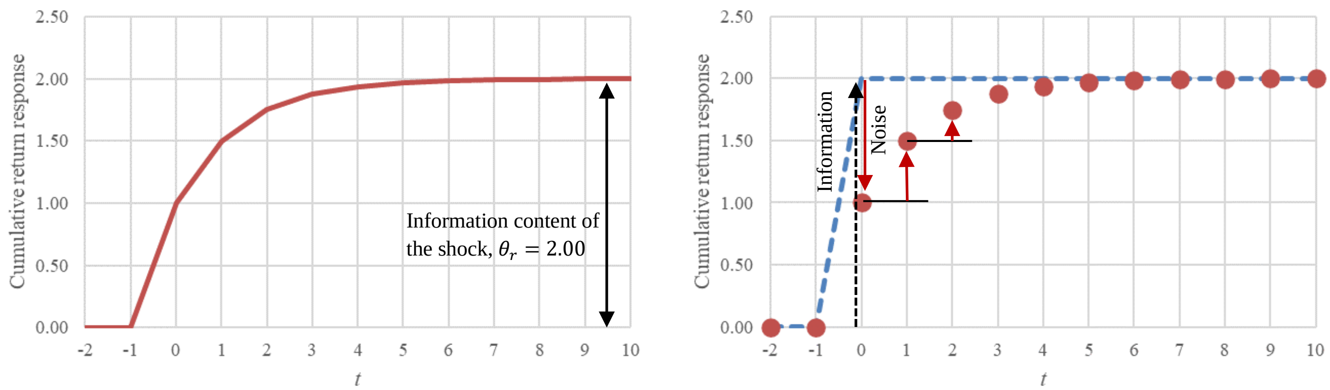
Vuolteenaho, T., 2002. What drives firm-level returns? Journal of Finance 57, 233–264.

Wurgler, J., 2000. Financial markets and the allocation of capital. Journal of Financial Economics 58, 187–214.

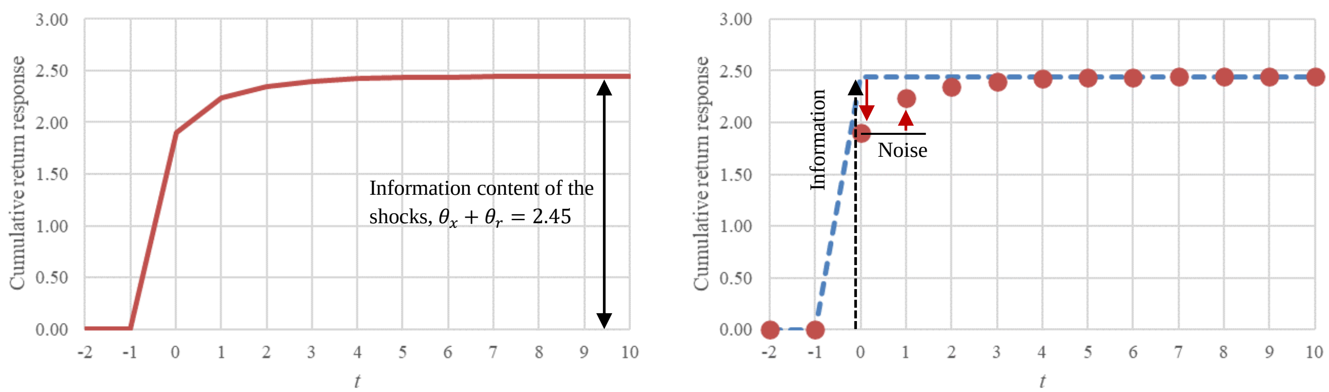
Panel A: Cumulative impulse response of returns to a shock to trading ($\varepsilon_{x,0} = 1$) showing over-reaction



Panel B: Cumulative impulse response of returns to a shock from non-trade sources ($\varepsilon_{r,0} = 1$) showing under-reaction



Panel C: Cumulative impulse response of returns to a shock to both trade and non-trade sources ($\varepsilon_{x,0} = \varepsilon_{r,0} = 1$)



(continued next page)

Panel D: Evolution of the estimated efficient price and actual prices for a series of shocks

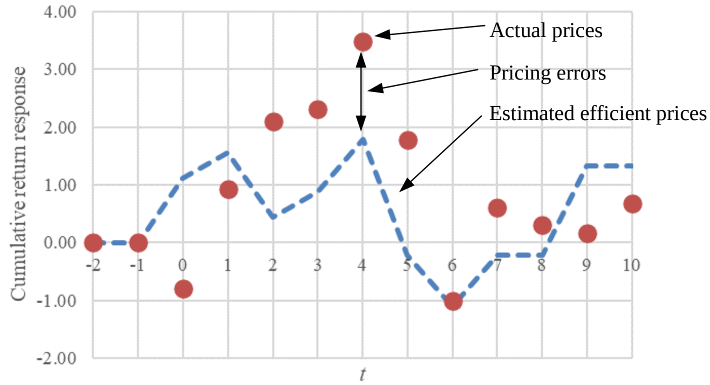


Figure 1. Example of separating information and noise in a Beveridge-Nelson decomposition.
The figure plots the cumulative impulse response functions of a simple illustrative VAR model given by Equations (2) in the paper. The vertical axis measures the estimated cumulative change in the log price of the stock as a result of the shock at time $t = 0$ with time on the horizontal axis. Panel A illustrates a unit shock to trading ($\varepsilon_{x,0} = 1$). Panel B illustrates a unit shock from non-trade sources ($\varepsilon_{r,0} = 1$). Panel C combines the two shocks ($\varepsilon_{x,0} = \varepsilon_{r,0} = 1$). The plots on the left-hand side are standard impulse response functions, while the plots on the right-hand side show how a realised return is broken up into information and noise components for the same shocks. Dots are prices (expressed as a cumulative percentage change from the price before the shock at $t = 0$), while the dashed line is the estimated efficient price corresponding to the cumulative information-driven returns. Finally, Panel D illustrates the estimated efficient price (dashed line) and actual prices (dots) for a hypothetical series of shocks: $\varepsilon_{x,t}$ shocks are (–2,1,2,1,2,0,–2,2,0,–1,0) and $\varepsilon_{r,t}$ shocks are (1,0,–1,0,0,–1,0,0,0,1,0).

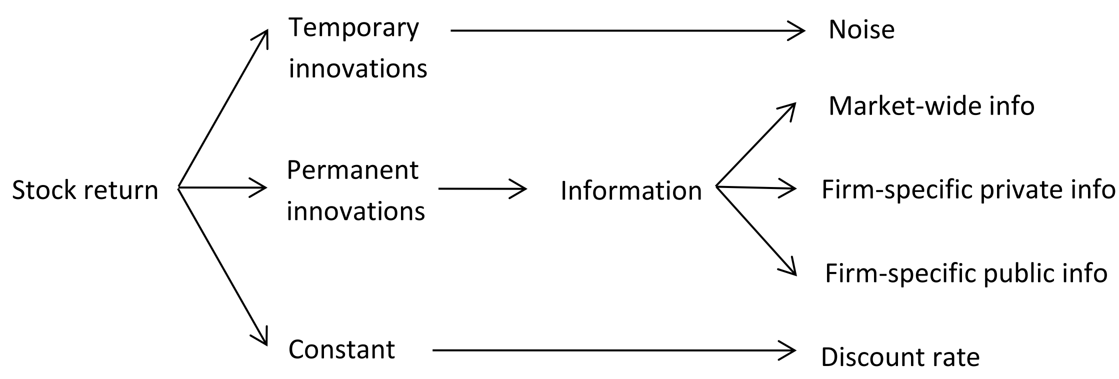
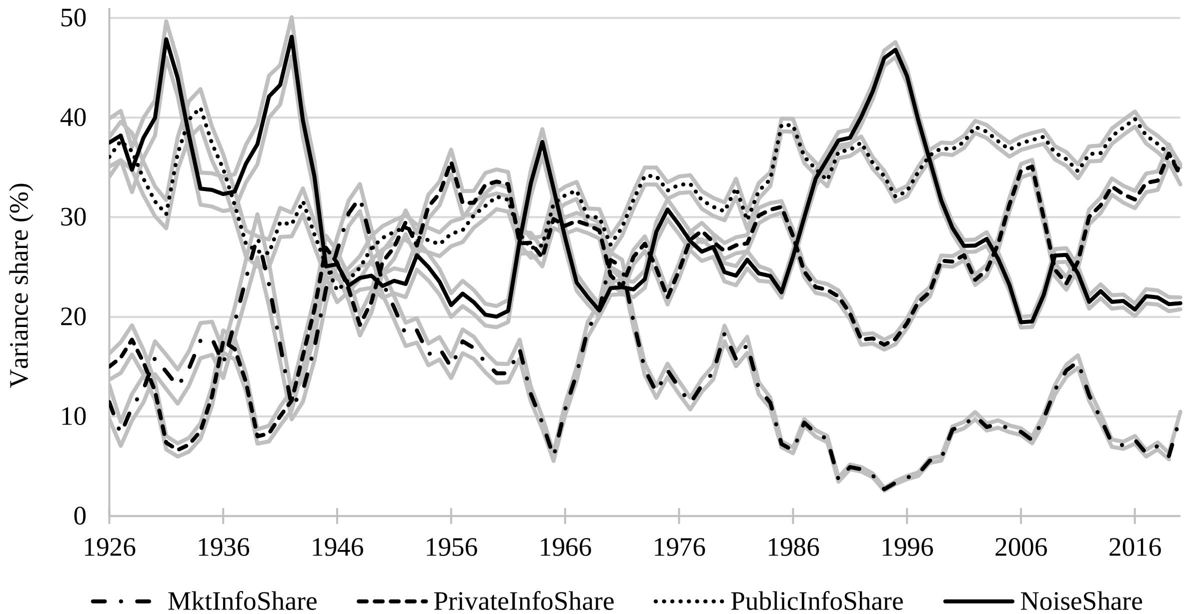


Figure 2. Stock return components in the baseline model.
In the baseline model stock returns are decomposed into temporary innovations (noise), three types of information (permanent innovations), and a constant (discount rate). The first four of these are the variance components in the baseline model, while the fifth (the discount rate) does not contribute to variance in the baseline model.

Panel A: Extended time period 1926-2020



Panel B: Subperiod 1960-2015

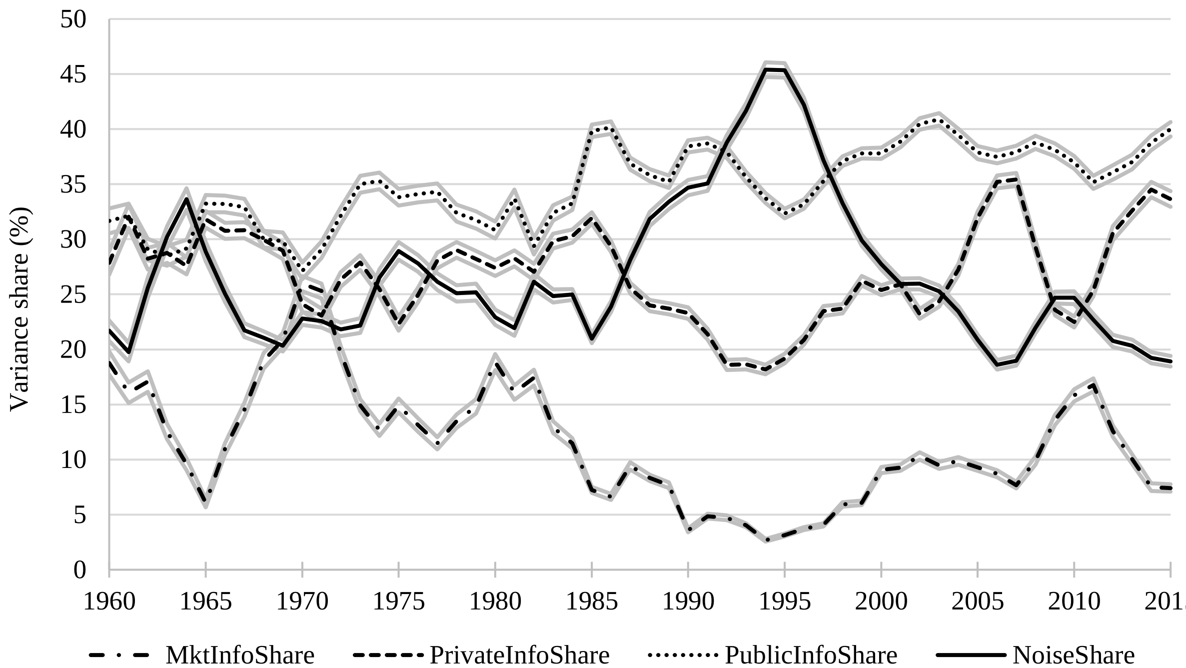


Figure 3. Stock return variance components for US stocks through time.

This figure shows the time-series trends in the percentage of stock return variance that is attributable to noise ($NoiseShare$), market-wide information ($MktInfoShare$), trading on private firm-specific information ($PrivateInfoShare$), and public firm-specific information ($PublicInfoShare$). The variance shares are calculated separately for each stock in each year based on a VAR model. For each of the variance shares in each year we then take a variance-weighted average across stocks. Light gray lines provide 99% confidence intervals. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ.

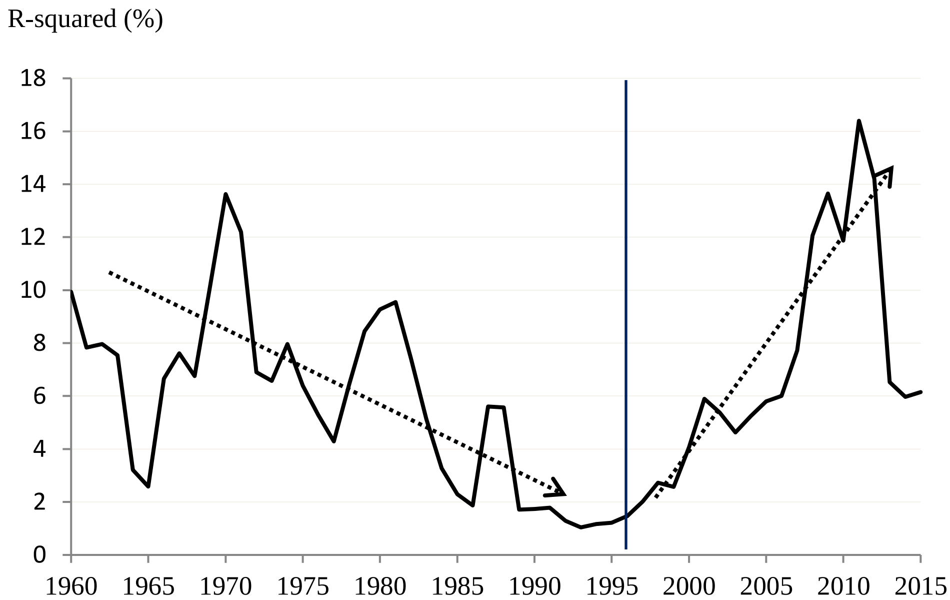


Figure 4. R^2 through time.
This figure shows the time-series trend in R^2 from 1960 to 2015. R^2 is calculated separately for each stock and each year by regressing individual daily stock returns on daily market returns, and then averaging across stocks. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ.

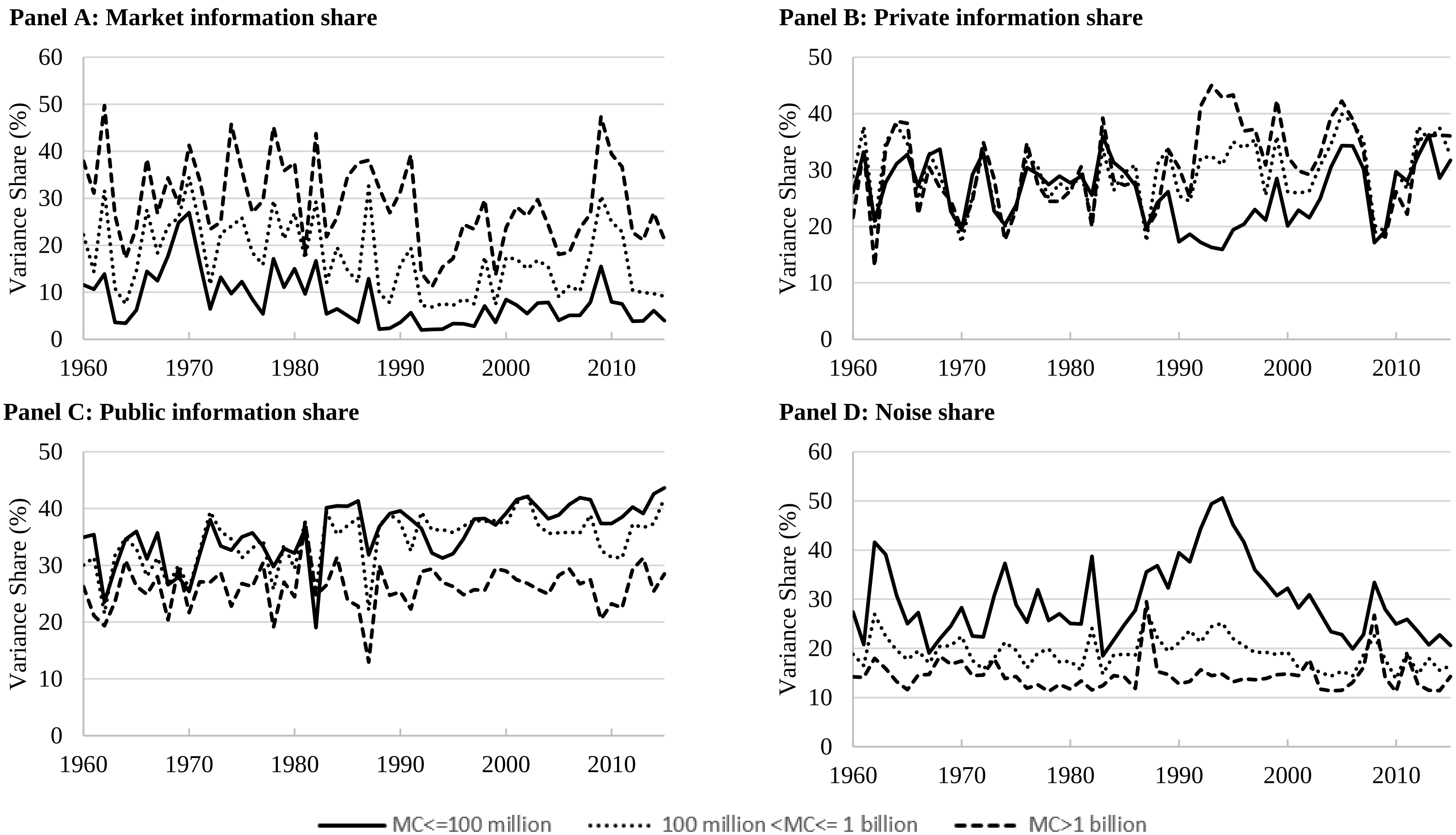


Figure 5. Variance components in size groups through time.
This figure shows the time-series trends in the percentage of stock return variance that is attributable to market-wide information (Panel A), private firm-specific information (Panel B), public firm-specific information (Panel C), and noise (Panel D) in three market capitalization groups: stocks with market capitalization less than \$100 million, market capitalization between \$100 million and \$1 billion, and market capitalization greater than \$1 billion. These breakpoints are in 2010 dollars and are adjusted for inflation forward and backward in time using the GDP price deflator. Each year stocks are assigned to one of the three groups based on their market capitalization at the start of the year. The variance component shares are calculated separately for each stock in each year and then averaged for each size group in each year. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015 (average of 4,362 stocks per year with a total of 22,025 stocks).

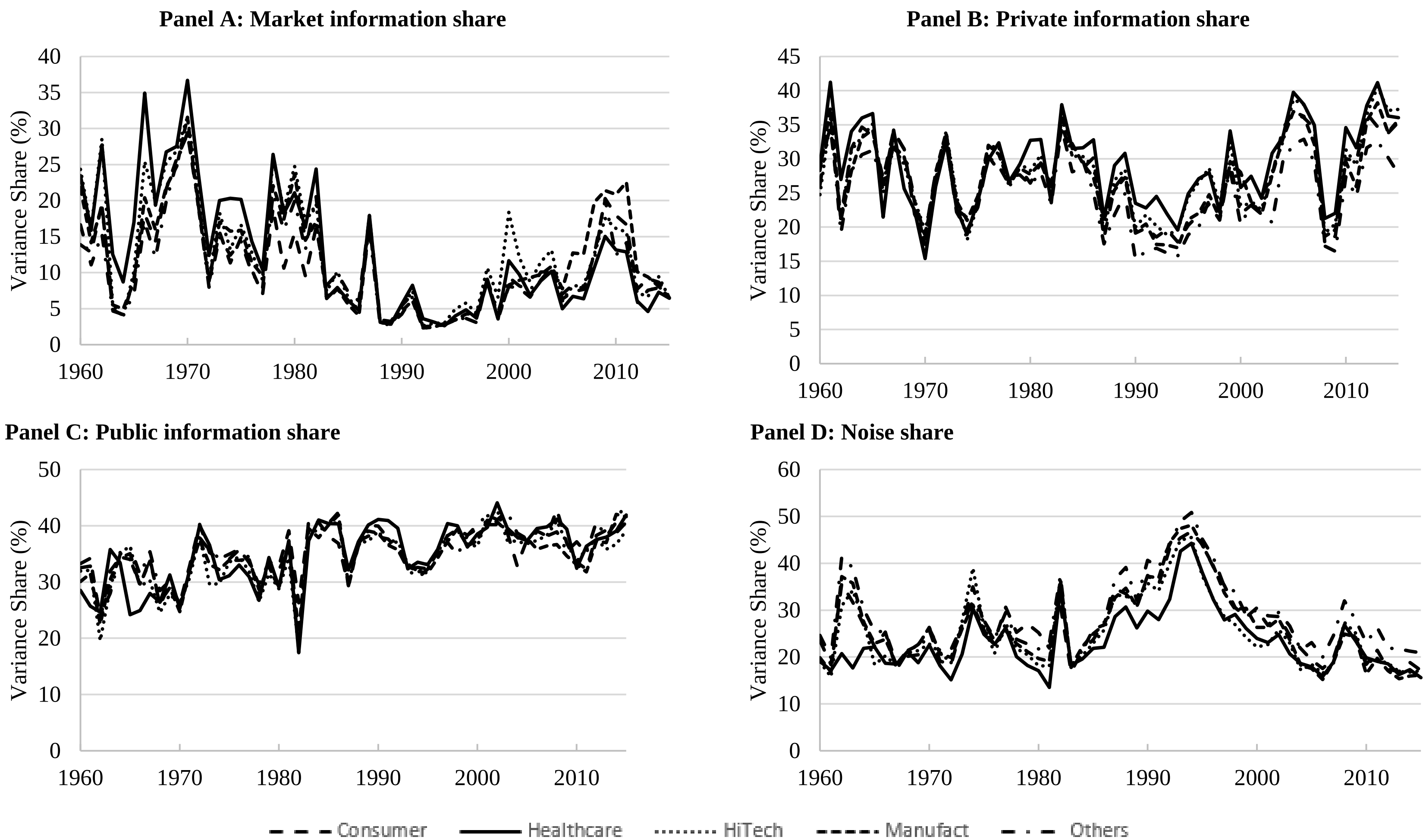


Figure 6. Variance components in major industry groups through time.
This figure shows the time-series trends in the percentage of stock return variance that is attributable to market-wide information (Panel A), private firm-specific information (Panel B), public firm-specific information (Panel C), and noise (Panel D) in five major industry groups. The *Consumer* group comprises the industries Consumer Durables, NonDurables, Wholesale, Retail, and some Services (Laundries, Repair Shops); the *Healthcare* group comprises the industries Healthcare, Medical Equipment, and Drugs; the *Manufact* group comprises the industries Manufacturing, Energy, and Utilities; the *HiTech* group comprises the industries Business Equipment, Telephone and Television Transmission; and the *Other* group comprises all other industries. The variance component shares are calculated separately for each stock in each year then averaged for each industry group in each year. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015 (average of 4,362 stocks per year with a total of 22,025 stocks).

Table 1. VAR coefficient estimates.

This table reports the mean coefficient estimates and the mean correlation between residuals for the baseline VAR model used to perform the variance decomposition. The VAR model is estimated separately for each stock in each year using daily observations. For the purpose of this table, each of the model coefficients is then averaged across stocks and years and reported in the table. Below each coefficient average, in parentheses, we report the percentage of negative statistically significant (at 5%) coefficients (first number in the parentheses) and the percentage of positive statistically significant (at 5%) coefficients (second number in the parentheses). The correlations column is computed similarly, but rather than reporting coefficients it reports the correlations of the residuals for pairs of variables in the VAR. The variables used in the VAR are: daily market returns in basis points ($r_{m,t}$), daily signed dollar volume in \$ thousands (x_t), and daily stock returns in basis points (r_t). The columns $l = 1$ to $l = 5$ correspond to lags of the independent variables. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015 (average of 4,362 stocks per year with a total of 22,025 stocks).

Dependent variable	Independent variable	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	Correlation
Panel A: Market return equation							
$r_{m,t}$	$r_{m,t-l}$	0.104	–0.032	0.018	–0.005	–0.010	
		(2.06%, 37.93%)	(8.13%, 0.34%)	(1.69%, 3.48%)	(5.05%, 1.81%)	(2.14%, 3.09%)	
	x_{t-l}	0.069	0.024	0.033	0.081	–0.034	0.134
		(2.36%, 2.83%)	(2.51%, 2.54%)	(2.41%, 2.48%)	(2.42%, 2.63%)	(2.40%, 2.44%)	(0.57%, 42.81%)
	r_{t-l}	–0.002	–0.0001	0.0006	0.0003	–0.001	0.229
		(3.76%, 3.07%)	(3.26%, 3.19%)	(3.04%, 3.14%)	(2.90%, 3.15%)	(3.22%, 2.83%)	(0.39%, 61.85%)
Panel B: Signed dollar volume equation							
x_t	$r_{m,t-l}$	–0.650	0.013	0.612	–0.529	–0.785	0.134
		(2.27%, 10.18%)	(2.42%, 2.91%)	(1.75%, 3.50%)	(2.05%, 2.85%)	(1.97%, 2.93%)	(0.57%, 42.81%)
	x_{t-l}	0.025	–0.016	0.003	–0.008	0.004	
		(8.69%, 14.74%)	(9.21%, 5.75%)	(5.45%, 6.14%)	(5.76%, 4.52%)	(4.26%, 5.15%)	
	r_{t-l}	1.736	0.631	0.196	–0.002	0.404	0.615
		(9.55%, 8.45%)	(4.66%, 4.83%)	(3.60%, 3.82%)	(3.01%, 3.50%)	(2.79%, 3.19%)	(0.05%, 99.42%)
Panel C: Stock return equation							
r_t	$r_{m,t-l}$	0.247	0.045	0.089	0.052	0.050	0.229
		(2.10%, 21.77%)	(2.93%, 4.78%)	(1.71%, 6.02%)	(2.46%, 4.37%)	(2.36%, 4.56%)	(0.39%, 61.85%)
	x_{t-l}	0.972	–0.077	–0.059	–0.150	–0.037	0.615
		(2.56%, 11.37%)	(3.06%, 3.97%)	(2.68%, 3.10%)	(2.80%, 2.66%)	(2.65%, 2.51%)	(0.05%, 99.42%)
	r_{t-l}	–0.112	–0.060	–0.030	–0.022	–0.007	
		(31.77%, 5.47%)	(16.40%, 2.61%)	(8.88%, 2.68%)	(6.27%, 2.58%)	(4.27%, 3.07%)	

Table 2. Stock return variance components in the baseline model.

This table reports the mean variance shares (expressed as percentages of variance) for the period from 1960 to 2015. Stock return variance is decomposed into market-wide information ($MktInfoShare$), private firm-specific information ($PrivateInfoShare$), public firm-specific information ($PublicInfoShare$), and noise ($NoiseShare$). Panel A reports full sample averages. Panel B splits the sample into two sub-periods from 1960 to 1996, and from 1997 to 2015. Panel C groups stocks into quartiles by size (market capitalization) with quartiles formed separately each year. Panel D groups stocks into major industry groups: the *Consumer* group comprises the industries Consumer Durables, NonDurables, Wholesale, Retail, and some Services (Laundries, Repair Shops); the *Healthcare* group comprises the industries Healthcare, Medical Equipment, and Drugs; the *Manufact* group comprises the industries Manufacturing, Energy, and Utilities; the *HiTech* group comprises the industries Business Equipment, Telephone and Television Transmission; and the *Other* group comprises all other industries. The variance component shares are calculated separately for each stock in each year then averaged across stocks within the corresponding quartile or group, using the variance-weighted mean. The 99% confidence intervals are in square brackets. We also report the differences in means for the post-1997 period minus the pre-1997 period (Panel B) and quartile 1 minus quartile 4 (Panel C) and corresponding t-statistics in parentheses. ***, **, and * indicate statistically significant differences at the 1%, 5%, and 10% levels using standard errors clustered by stock and by year. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ (average of 4,362 stocks per year with a total of 22,025 stocks).

	$MktInfoShare$ (%)	$PrivateInfoShare$ (%)	$PublicInfoShare$ (%)	$NoiseShare$ (%)
Panel A: Full sample				
	8.62	24.00	36.67	30.71
	[8.56–8.67]	[23.92–24.08]	[36.59–36.76]	[30.61–30.81]
Panel B: Sub-periods				
1960–1996	7.11	22.61	34.86	35.42
	[7.04–7.19]	[22.50–22.72]	[34.75–34.97]	[35.28–35.56]
1997–2015	10.19	25.46	38.58	25.78
	[10.10–10.28]	[25.33–25.58]	[38.45–38.71]	[25.65–25.91]
Difference (Post-Pre 1997)	3.07	2.85	3.72	–9.64
	(1.84)*	(1.83)*	(3.74)***	(–4.85)***
Panel C: Quartiles by size (market capitalization)				
Q1=low	5.58	21.85	37.01	35.56
	[5.51–5.66]	[21.70–22.00]	[36.85–37.16]	[35.37–35.76]
Q2	8.84	25.04	37.43	28.69
	[8.73–8.95]	[24.87–25.21]	[37.24–37.62]	[28.48–28.90]
Q3	14.21	27.90	36.80	21.08
	[14.07–14.36]	[27.72–28.08]	[36.62–36.99]	[20.91–21.25]
Q4=high	21.09	30.33	31.67	16.92
	[20.92–21.26]	[30.14–30.51]	[31.507–31.84]	[16.79–17.04]
Difference (Q1–Q4)	–15.51	–8.48	5.34	18.64
	(–14.82)***	(–6.15)***	(5.79)***	(11.08)***
Panel D: Industry groups				
Consumer	7.98	23.72	36.37	31.93
	[7.85–8.10]	[23.53–23.90]	[36.18–36.56]	[31.71–32.16]
Healthcare	7.67	27.01	38.01	27.30
	[7.49–7.85]	[26.71–27.31]	[37.70–38.33]	[27.00–27.64]
HiTech	9.50	25.26	37.12	28.13
	[9.36–9.63]	[25.07–25.44]	[36.92–37.31]	[27.90–28.35]
Manufact	10.03	24.28	35.29	30.40
	[9.89–10.17]	[24.11–24.45]	[35.12–35.47]	[30.18–30.61]
Other	7.75	21.89	36.80	33.56
	[7.65–7.86]	[21.74–22.04]	[36.65–36.96]	[33.37–33.75]

Table 3. Determinants of stock return variance shares.

This table reports the results from panel regressions of stock-year observations in which the dependent variables are the shares of stock return variance attributable to market-wide information ($MktInfoShare_{i,t}$), private firm-specific information ($PrivateInfoShare_{i,t}$), public firm-specific information ($PublicInfoShare_{i,t}$), and noise ($NoiseShare_{i,t}$). For comparison, we also include regressions in which the dependent variable is the $(1 - R^2)$ measure of firm-specific information. The explanatory variables in Panel A are as follows. D_t^{POST} is a dummy variable that takes the value of one after 1997 and zero before. $lnP_{i,t}$ is the log price and $lnMC_{i,t}$ is the log market capitalization. $D_i^{Consumer}$, $D_i^{Healthcare}$, D_i^{HiTech}, and $D_i^{Manufact}$ are dummy variables that indicate the firm's industry group (the Other Industry grouping is the omitted category). $VarianceRatio_{i,t}$ is a measure of the amount of information impounded in prices during the trading session (volatility of open-to-close returns divided by the volatility of overnight close-to-open returns). $Delay_{i,t}$ is a measure of the delay with which a stock's prices respond to market-wide information. In Panel B we add an extended set of stock characteristics including the following. $InstoHold_{i,t}$ is the percentage of outstanding shares held by institutional investors. $HardToValue_{i,t}$ is a proxy for hard to value stocks, estimated as the first principal component of share turnover, idiosyncratic volatility, and firm age (following Kumar, 2009). $ILLIQ_{i,t}$ is the Amihud (2002) measure of illiquidity. $BidAskSpread_{i,t}$ is the average effective bid-ask spread. $Leverage_{i,t}$ is (long-term debt + current liabilities)/(long-term debt + current liabilities + shareholder's equity). The sample includes stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015. t-statistics are in parentheses using standard errors clustered by stock and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

Panel A: Basic stock characteristics as determinants

Variable	$MktInfoShare_{i,t}$	$PrivateInfoShare_{i,t}$	$PublicInfoShare_{i,t}$	$NoiseShare_{i,t}$	$(1 - R^2)$
Intercept	19.38	12.51	31.66	36.45	78.59
	(7.03)***	(6.49)***	(18.80)***	(16.17)***	(34.12)***
D_t^{POST}	0.97	–1.37	4.29	–3.88	–2.30
	(1.03)	(–1.34)	(5.60)***	(–4.01)***	(–3.34)***
$lnP_{i,t}$	0.01	–1.02	0.21	0.80	0.57
	(0.02)	(–1.95)*	(0.68)	(2.06)**	(1.97)**
$lnMC_{i,t}$	0.60	2.52	–1.02	–2.09	–1.38
	(2.01)**	(8.92)***	(–4.78)***	(–7.35)***	(–6.15)***
$D_i^{Consumer}$	–0.47	4.40	0.21	–4.13	1.36
	(–1.63)	(12.70)***	(0.53)	(–9.10)***	(5.82)***
$D_i^{Healthcare}$	–1.07	6.64	0.72	–6.29	2.99
	(–2.15)**	(14.73)***	(1.39)	(–13.43)***	(6.25)***
D_i^{HiTech}	0.58	4.77	–0.07	–5.27	1.80
	(0.91)	(7.42)***	(–0.16)	(–10.58)***	(4.58)***
$D_i^{Manufact}$	0.99	2.75	–0.16	–3.58	–0.29
	(1.59)	(7.43)***	(–0.31)	(–7.93)***	(–0.72)
$VarianceRatio_{i,t}$	–0.05	0.59	0.84	–1.39	0.44
	(–0.42)	(3.72)***	(7.17)***	(–5.91)***	(2.54)**
$Delay_{i,t}$	–21.24	2.46	6.91	11.87	29.43
	(–8.97)***	(1.45)	(5.93)***	(8.58)***	(10.56)***
Observations	137,903	137,903	137,903	137,903	137,903
R^2	41.1%	9.8%	5.1%	31.3%	64.8%

Panel B: Extended set of stock characteristics as determinants					
Variable	$MktInfoShare_{i,t}$	$PrivateInfoShare_{i,t}$	$PublicInfoShare_{i,t}$	$NoiseShare_{i,t}$	$(1-R^2)$
Intercept	17.32	13.92	36.25	32.52	80.17
	(5.97)***	(5.28)***	(17.51)***	(12.46)***	(27.47)***
D_t^{POST}	1.78	–2.59	3.26	–2.44	–3.07
	(2.63)***	(–2.36)**	(5.17)***	(–2.18)**	(–4.60)***
$lnP_{i,t}$	–0.73	–1.39	0.033	2.09	0.71
	(–1.32)	(–2.15)**	(0.11)	(6.28)***	(1.95)*
$lnMC_{i,t}$	1.07	2.27	–1.39	–1.95	–1.49
	(5.91)***	(8.95)***	(–10.16)***	(–8.69)***	(–6.68)***
$D_i^{Consumer}$	–1.10	4.97	–0.10	–3.77	2.36
	(–2.88)***	(11.85)***	(–0.23)	(–7.34)***	(7.05)***
$D_i^{Healthcare}$	–1.62	6.82	0.64	–5.83	3.78
	(–2.51)**	(14.02)***	(1.08)	(–10.54)***	(6.31)***
D_i^{HiTech}	0.29	5.23	–0.43	–5.08	2.37
	(0.38)	(7.14)***	(–0.83)	(–9.09)***	(4.60)***
$D_i^{Manufact}$	0.75	3.19	–0.26	–3.67	0.19
	(1.23)	(7.41)***	(–0.51)	(–6.67)***	(0.39)
$VarianceRatio_{i,t}$	–0.094	0.52	0.84	–1.26	0.52
	(–0.68)	(3.17)***	(7.00)***	(–6.14)***	(2.46)**
$Delay_{i,t}$	–21.28	5.61	8.60	7.07	31.28
	(–8.60)***	(2.78)***	(7.14)***	(4.22)***	(9.99)***
$InstoHold_{i,t}$	0.014	0.040	–0.013	–0.041	–0.04
	(0.96)	(2.32)**	(–1.37)	(–4.78)***	(–2.56)**
$HardToValue_{i,t}$	–0.30	1.88	0.55	–2.13	1.27
	(–1.03)	(5.36)***	(2.38)**	(–6.83)***	(4.58)***
$ILLIQ_{i,t}$	0.003	–0.014	–0.015	0.026	–0.012
	(1.88)*	(–2.84)***	(–2.24)**	(2.56)**	(–2.67)***
$BidAskSpread_{i,t}$	0.013	–0.068	–0.052	0.110	–0.032
	(1.81)*	(–3.45)***	(–2.84)***	(3.19)***	(–2.89)***
$Leverage_{i,t}$	0.012	–0.083	0.094	–0.022	–0.013
	(0.67)	(–2.32)**	(2.89)***	(–0.64)	(–0.72)
Observations	82,116	82,116	82,116	82,116	82,116
R^2	42.21%	11.74%	8.37%	34.15%	66.24%

Table 4. Effects of the tick size changes on noise and information.

This table reports the results from panel regressions of stock-year observations in which the dependent variable is given in the column heading: $PrivateInfo_{i,t}$ is the amount of stock return variance attributed to private firm-specific information, $Noise_{i,t}$ is the amount of stock return variance attributed to noise, and $NoiseShare_{i,t}$ is the share (percentage) of variance attributed to noise. The models examine the effects of tick size reductions from eighths of a dollar to sixteens of a dollar on June 24, 1997 using two years of data around the change (1996, 1998). D_t^{POST} is a dummy variable that takes the value of one after 1997 and zero before. $Q1_i$, $Q2_i$, and $Q3_i$ are dummy variables that indicate the price quartile to which the firm belongs (the highest price quartile, $Q4_i$, is the omitted category). $lnP_{i,t}$ is the log price. t-statistics are in parentheses using standard errors clustered by stock. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes all stocks listed on NYSE, AMEX, and NASDAQ.

Variable	Model 1 $PrivateInfo_{i,t}$	Model 2 $PrivateInfo_{i,t}$	Model 3 $Noise_{i,t}$	Model 4 $Noise_{i,t}$	Model 5 $NoiseShare_{i,t}$	Model 6 $NoiseShare_{i,t}$
Intercept	1.39	3.32	1.02	4.74	19.87	48.66
	(64.83)***	(85.95)***	(66.15)***	(98.32)***	(53.31)***	(77.12)***
D_t^{POST}	0.15	0.58	0.25	0.29	–0.76	–11.63
	(5.86)***	(11.12)***	(12.86)***	(5.31)***	(–1.98)**	(–17.01)***
$D_t^{POST} \times Q1_i$	0.43		0.28		–5.55	
	(8.83)***		(5.58)***		(–8.36)***	
$D_t^{POST} \times Q2_i$	0.16		0.07		–5.87	
	(3.54)***		(1.93)*		(–8.80)***	
$D_t^{POST} \times Q3_i$	0.08		0.04		–3.56	
	(1.72)*		(1.20)		(–5.14)***	
$Q1_i$	1.26		2.48		18.87	
	(36.53)***		(64.38)***		(31.69)***	
$Q2_i$	0.57		0.92		10.38	
	(17.03)***		(31.80)***		(16.55)***	
$Q3_i$	0.21		0.38		7.23	
	(6.58)***		(15.64)***		(11.49)***	
$D_t^{POST} \times lnP_{i,t}$		–0.11		0.02		2.77
		(–6.04)***		(0.80)		(11.80)***
$lnP_{i,t}$		–0.56		–1.09		–7.77
		(–40.45)***		(–63.33)***		(–34.67)***
Observations	15,440	15,440	15,440	15,440	15,440	15,440
R^2	20.3%	22.3%	45.8%	49.3%	11.0%	11.7%

Table 5. Effect of analyst coverage on variance components.

This table reports the results from difference-in-differences regressions in which we examine the causal effect of an exogenous drop in analyst coverage (due to brokerage mergers and closures) on variance components. The dependent variables are shares of stock return variance attributable to market-wide information ($MktInfoShare_{i,t}$), private firm-specific information ($PrivateInfoShare_{i,t}$), public firm-specific information ($PublicInfoShare_{i,t}$), and noise ($NoiseShare_{i,t}$). The independent variable of interest is $CoverageShock_{i,t}$ which is the number of broker disappearances due to mergers and closures of brokerage houses during the past two years (max of the $t-1$ and $t-2$ values). The regressions include stock and time (year) fixed effects. t-statistics are in parentheses using standard errors clustered by stock and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on NYSE, AMEX, and NASDAQ from 1987 to 2011 (the period containing the analyst coverage shock events).

Variable	$MktInfoShare_{i,t}$	$PrivateInfoShare_{i,t}$	$PublicInfoShare_{i,t}$	$NoiseShare_{i,t}$
Intercept	7.52	–3.89	–4.71	1.08
	(112.75)***	(–66.40)***	(–102.49)***	(34.07)***
$CoverageShock_{i,t}$	0.31	0.45	–1.47	0.71
	(0.43)	(0.70)	(–2.96)***	(2.09)**
Observations	116,879	116,879	116,879	116,879
R^2	23.0%	7.6%	2.5%	5.4%
Fixed Effects	Stock, Time	Stock, Time	Stock, Time	Stock, Time

Table 6. Effect of mutual fund fire sales and discretionary trade shocks on variance components.

This table reports the results from difference-in-differences regressions of stock-quarter observations in which we examine the causal effect of an exogenous shock to mutual fund trading of an underlying stock, driven by investor redemptions or inflows. The dependent variables are shares of stock return variance attributable to market-wide information ($MktInfoShare_{i,t}$), private firm-specific information ($PrivateInfoShare_{i,t}$), public firm-specific information ($PublicInfoShare_{i,t}$), and noise ($NoiseShare_{i,t}$). $DiscretionaryTradeShock_{i,t}$ takes the value of one for stock-quarters that experience strong discretionary buying or selling by mutual funds in response to a redemption shock (a proxy for informed mutual fund trading, defined as the bottom and top deciles of mutual fund discretionary trading). Similarly, $ExpectedTradeShock_{i,t}$ takes the value of one for stock-quarters that experience strong non-discretionary fire sales by mutual funds due to fund redemptions (a proxy for uninformed mutual fund trading, defined as the bottom decile of mutual fund non-discretionary trading). The regressions include stock and time (year-quarter) fixed effects. t-statistics are in parentheses using standard errors clustered by stock and year-quarter. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on NYSE, AMEX, and NASDAQ from 1980 to 2015 (the period containing the mutual fund trading shock events).

Variable	$MktInfoShare_{i,t}$	$PrivateInfoShare_{i,t}$	$PublicInfoShare_{i,t}$	$NoiseShare_{i,t}$
Intercept	–0.50	1.93	0.40	–1.83
	(–7.49)***	(12.08)***	(2.85)***	(–23.20)***
$DiscretionaryTradeShock_{i,t}$	0.33	0.50	0.16	–0.98
	(3.44)***	(4.71)***	(1.59)	(–12.13)***
$ExpectedTradeShock_{i,t}$	–0.23	0.19	0.07	–0.03
	(–1.68)*	(1.39)	(0.54)	(–0.32)
Observations	470,307	470,307	470,307	470,307
R^2	27.1%	10.0%	5.8%	7.1%
Fixed Effects	Stock, Time	Stock, Time	Stock, Time	Stock, Time