January 18, 2022

A strategic trade optimization approach to identifying private information

Dion Bongaerts, Dominik Rösch, and Mathijs van Dijk*

Abstract

We propose a strategic trade optimization approach to identifying private information. In our model, investors are exposed to liquidity and private information shocks and strategically optimize their trading across stocks taking into account price impact (Kyle's λ). The model yields a simple private information measure: $\lambda \times OIB$ (order imbalance). Intuitively, observed order imbalance is more likely to be information-driven when trading is expensive. Consistent with our measure capturing private information, we show empirically that it is greater for smaller firms with higher analyst dispersion, helps explain return reversals, predicts return volatility, and increases before M&A announcements and after analyst coverage terminations.

*Bongaerts and van Dijk are at the Rotterdam School of Management, Erasmus University; Rösch is at the State University of New York, Buffalo. E-mail addresses: dbongaerts@rsm.nl, drosch@buffalo.edu, and madijk@rsm.nl. We thank Yakov Amihud, Tarun Chordia, Mathijs Cosemans, Frank de Jong, Dave Denis, Andrei Dubovik, Cam Harvey, Joel Hasbrouck, Sahn-Wook Huh, Andrei Kirilenko, Andy Koch, Pete Kyle, Francis Longstaff, Anthony Lynch, Frédéric Malherbe, Stefan Obernberger, Franz Palm, Marios Panayides, Ioanid Rosu, Zacharias Sautner, Frederik Schlingemann, Vincent van Kervel, Stijn Van Nieuwerburgh, Dimitri Vayanos, Bart Zhou Yueshen, Darya Yuferova, and seminar participants at Cheung Kong Graduate School of Business, Erasmus University, HEC Lausanne, University of Pittsburgh, the Conference on Financial Econometrics and Empirical Asset Pricing (Lancaster, U.K.), the Maastricht Workshop on Advances in Quantitative Economics (Maastricht, the Netherlands), and the 2nd CEPR-Imperial-Plato Market Innovator (MI3) Conference on Market Structure in Europe and Beyond (London, U.K.) for valuable comments. We thank Santanu Kundu and Yihe Yu for excellent research assistance. This work was carried out on the National e-infrastructure with the support of SURF Foundation. We thank SURFsara for technical support on computing and storage, and OneMarketData for the use of their OneTick software. Van Dijk gratefully acknowledges financial support from the Vereniging Trustfonds Erasmus Universiteit Rotterdam and from the Netherlands Organisation for Scientific Research through a "Vidi" grant.

1. Introduction

The notion of private information plays an important role in many theoretical models of market microstructure, asset pricing, and corporate finance. Such models show, for example, that informed trading affects price efficiency and market liquidity, the cost of capital, and the incentives for corporate investments.[1] However, measuring private information and informed trading empirically remains a considerable challenge.

We propose a new private information measure for individual securities based on a theoretical strategic trade optimization model. For cross-sectional applications, the measure is equal to the product of the security's price impact parameter (Kyle's λ) and its aggregate order imbalance (OIB): $\lambda \times OIB$. For general applications, a correction term for aggregate uninformed trading, given by the $\lambda \times OIB$ of a benchmark security that is not subject to informed trading, needs to be subtracted.[2] The measure is straightforward to estimate for any security over any time period. It also has a clear intuition: investors only trade securities that are expensive to trade when they have valuable private information, so order imbalance is more likely to be information-driven when price impact is high.

We set up a one-period model, in which M investors optimize their trading across N securities when hit by investor-specific liquidity shocks and private information shocks. Investors maximize speculative profits from trading on information, while minimizing transaction costs (stemming from security-specific linear price impact λ), and satisfying the budget constraint implied by the liquidity shock. In contrast to Kyle (1985), investors in our model are "strategic" in the sense that they optimize their uninformed trading in the cross-section of securities based on price impact. In particular, when hit by a

[1] See, among many others, Grossman and Stiglitz (1980), Glosten and Milgrom (1985), Kyle (1985), Fishman and Hagerty (1989), Manove (1989), Easley, Hvidkjaer, and O'Hara (2002), Dow and Rahi (2003), Easley and O'Hara (2004), Goldstein and Guembel (2008), and Edmans (2009).

[2] This correction term is common to all securities and thus drops out in cross-sectional applications.

liquidity shock, investors optimally spread their trading over all securities, such that the marginal transaction costs for all securities are equal. As a result, liquidity-motivated trading is inversely proportional to the price impact parameter λ, and more uninformed trading is done in more liquid securities.

When receiving a private information shock about a security, investors trade-off speculative profits against transaction costs. They optimally engage in speculative trading in that security with an amount that is inversely related to its λ; this behavior aligns with that of the informed trader in Kyle (1985, see his Eq. (2.6)). Furthermore, investors trade other securities in the opposite direction to finance the speculative trade. As with liquidity-motivated trades, such "funding" trades are optimally spread across securities, where again the amount of trading in each security is inversely proportional to its λ.[3]

The first main result of our model is thus that liquidity-motivated and funding order flow of a given investor are distributed across securities in the same way (i.e., inversely proportional to λ) – although liquidity-motivated and funding trades could be in opposite directions. Since λ is security-specific but not investor-specific, this result also obtains for the aggregate order flow across all investors, which implies that all aggregate uninformed (that is, liquidity-motivated and funding) order flow in the market can be summarized as if it were generated by one implicit market-aggregate liquidity shock.

This result reduces the dimensionality of the signal extraction problem and allows us to uncover the private information shock for any security from its aggregate order flow (aggregated across all investors). We derive a system of equations that describes the aggregate order flows in all securities as functions of the price impact parameters

[3]In our baseline model specification, we assume that for each security there is only a single investor with a non-zero information shock, and that this information shock is perfectly accurate. Internet Appendix IA.2 shows that our main result also obtains with multiple informed investors and/or with noisy information signals. Furthermore, Internet Appendix IA.2 shows that, under mild conditions, the model yields similar results when we introduce a liquid facility (such as a savings account or credit line) that some investors can use to finance speculative and liquidity-motivated trades.

and the private information shocks of all securities, as well as the market-aggregate liquidity shock. Solving the system yields a closed-form expression for each security's private information shock as a function of the cross-section of price impact parameters and aggregate order flows – the latter of which can be empirically estimated using order imbalances (OIB or the volume of buyer- minus seller-initiated trades).

However, if information shocks can materialize for all securities, this system of equations is underidentified, since we need to extract N security-specific private information shocks plus a market-aggregate liquidity shock from the aggregate order flow of N securities. To resolve the underidentification, we assume that there is one security that is not subject to private information shocks. The order flow in this information-free benchmark security stems only from uninformed (liquidity-motivated and funding) trades, which are distributed identically across securities and are inversely proportional to λ. Thus, the product of the benchmark security's λ and its order flow is a sufficient statistic to summarize all uninformed order flow. As the second main result of the model, we thereby obtain a remarkably simple closed-form expression for a security's private information shock: the product of the security's λ and its order imbalance OIB minus the product of the benchmark security's λ and OIB.

The benchmark's term in this expression serves as a way of "correcting" the aggregate order flow in a security for uninformed order flow. However, when comparing private information in the cross-section of securities (which is the focus of our empirical analyses), the benchmark's term drops out since it is common to all individual securities over a given time period. The expression for an individual security's private information shock then collapses into a very simple measure of private information: $\lambda \times OIB$. Our third main result is thus that the $\lambda \times OIB$ of a given security can be directly used in cross-sectional applications without any correction for uninformed order flow. This measure has an intuitive revealed preference interpretation: trading in securities that are more expensive to trade is more likely to be information-driven.[4]

Our private information measure $\lambda \times OIB$ is reminiscent of Kyle's (1985) result – in a single-security setting – that private information can be summarized by price impact times informed order flow. However, Kyle's expression cannot be used empirically since observed order flow is comprised of both informed and uninformed flow. We get around this problem by adopting a multi-security setting in which, rather than being myopic, uninformed trading is strategically spread across securities based on their price impact.[5] Our trade optimization approach allows us to estimate private information based on price impact times *total* order flow (i.e., the sum of informed and uninformed order flow, which is observed) and shows that, for cross-sectional applications, there is no need for a correction for uninformed order flow.

In our empirical implementation, we estimate our private information measure for all NYSE stocks each day over 2001-2014 using intraday price and transaction data (based on 18.6 billion transactions in total) – although it could also be estimated at higher or lower frequencies. We estimate daily λ based on intraday data following Goyenko, Holden, and Trzcinka (2009) and Hasbrouck (2009) and daily OIB by signing individual trades using the Lee and Ready (1991) algorithm. Our final sample consists of 1,338 NYSE-listed stocks over 2001-2014. As the benchmark security, we use the SPDR S&P500 ETF

[4]Although not used it in the paper, our model also allows for the extraction of the three components of a security's aggregate order flow: (i) liquidity-motivated, (ii) speculative, and (iii) funding order flow. It is thus not only possible to obtain closed-form expressions for the (signed) *private information shock* underlying the informed trading in individual securities, but also for the *amount of informed trading* – which can be useful for other applications and can be implemented using estimates of λ and OIB.

[5]We note that strategic uninformed trading in the cross-section has implications for equilibrium price impact parameters if derived endogenously. Their expressions become far more complicated than in Kyle (1985), because they are interdependent and need to be solved jointly across securities. In Internet Appendix IA.3, we show that the resulting expressions are intractable but can be derived and solved (under some conditions). Hence, under these conditions, our assumption of linear price impact is consistent with equilibrium behavior of investors and market makers in a setting with endogenous price impact. Therefore, endogenizing price impact does *not* affect the main results of our model. In particular, the simple expression for our private information measure is not affected.

(ticker "SPY"). We argue that the SPDR is a reasonable benchmark security since it is highly traded, since the scope for market-wide private information is arguably limited (e.g., Baker and Stein, 2004), and since the SPDR is unlikely to be used for trading on private information of individual securities.

We present six pieces of evidence that support the use of our measure as a private information proxy. First, we show that our private information measure (which can assume both positive and negative values since information shocks can be positive or negative) tends to be larger in absolute value for smaller firms, firms with higher analyst dispersion, higher trading volume, and higher price impact parameters – consistent with the notion that there is more scope for private information for smaller firms with greater uncertainty and more asymmetric information.

Second, we find that the cross-section of daily stock returns is positively and significantly related to the contemporaneous simplified private information measure $\lambda \times OIB$, even after controlling for λ and OIB separately. This finding is consistent with the idea that stocks with a more positive (negative) private information shock on a given day have a more positive (negative) realized stock return on that day.

Third, if our measure indeed picks up private information, return reversals should be weaker following larger positive or negative realizations of our measure. After all, the price impact of informed order flow should be permanent, while the price impact of uninformed order flow should be transient (e.g., Kyle, 1985; Admati and Pfleiderer, 1988; Glosten and Harris, 1988; Sadka, 2006). Consistent with this conjecture, we find that returns revert significantly less following stock-days with larger absolute values of the simplified private information measure.

Fourth, we find evidence consistent with the prediction of market microstructure models that informed trading should intensify ahead of the arrival of news. In particular, we show that firms with larger absolute values for the simplified private information measure tend to experience greater return volatility on the next day, even after controlling for lagged returns, lagged volatility, and λ and OIB separately.

Fifth, we examine private information around merger and acquisition (M&A) announcements. We focus on M&A transactions with publicly traded targets with a significantly positive price run-up in the ten days before the announcement, which is suggestive of pre-announcement incorporation of private information into the stock price. Consistent with informed trading in the target's stock ahead of the M&A announcement, we show that our simplified private information measure has large positive values before the announcement, and values close to zero after the announcement. We also find a strong cross-sectional relation between $\lambda \times OIB$ and the strength of the pre-announcement price run-up across the different M&A transactions.

Sixth, following Kelly and Ljungqvist (2012), we use the termination of analyst coverage of individual stocks due to a number of U.S. brokerage firms closing their research departments as an exogenous shock to public information production. In a difference-in-differences (DiD) analysis, we find that our private information measure increased for stocks affected by this shock, in line with notion that the loss of analyst coverage increased asymmetric information and the prospect for trading based on private information.

Our main contribution is a new, simple measure of private information that can be readily applied to any security over any time period. Our measure combines two well-known notions (λ and OIB) that have thus far been used individually to gauge a security's degree of asymmetric information. Kyle's (1985) λ is a widely accepted ex ante proxy for asymmetric information (e.g., Glosten and Harris, 1988; Hasbrouck, 1991). However, it is less well-suited as an ex post proxy for the amount of informed trading or the realized private information shock underlying informed trading. Other studies (e.g., Easley, Engle, O'Hara, and Wu, 2008; Kaul and Stoffman, 2008; Holden and Jacobsen, 2014) use absolute OIB as an ex post measure of informed trading. However, OIB

can also indicate liquidity-motivated trading, especially when price impact is low. Our $\lambda \times OIB$ measure encompasses these two notions and reflects the intuition that a private information measure should take into account both the amount and the costs of trading.

We complement the large literature on the popular "probability of informed trading" (PIN) measure developed by Easley, Kiefer, O'Hara, and Paperman (1996) and Easley et al. (2002), which is based on observed order flow alone and disregards price impact.[6] To the best of our knowledge, we are the first to use a strategic trade optimization model instead of a structural microstructure model to separate informed trading from liquidity-motivated trading. We are also the first to propose a way to measure the *magnitude* and *direction* of the private information shock underlying the trading in an individual security over a given period. An advantage of our approach relative to PIN is that it does not require estimating a structural model based on a long time-series of transaction data for individual securities (and can thus be estimated even at high frequencies). We thereby circumvent the numerical overflow problems (Holden and Jacobsen, 2014, p. 1757) and other potential drawbacks of PIN such as its counterintuitive behavior around corporate news announcements (see footnote 2 of Back, Crotty, and Li (2017)). To distinguish $\lambda \times OIB$ from PIN, we control for the daily PIN measures proposed by Brennan et al. (2018) in our empirical analyses.

We further complement studies that develop alternative measures of the degree of informed trading, such as Johnson and So (2018), who propose a measure of informed trading based on abnormal volume imbalances across stock and options markets, and Back et al. (2017), who develop a hybrid of the PIN and Kyle models and also identify informed trading based on order flow and price impact – although the channel is very different. Their price impact process contains information because it is endogenous, while

[6] Papers refining this approach include Odders-White and Ready (2008), Duarte and Young (2009), Easley, López de Prado, and O'Hara (2011, 2012), and Brennan, Huh, and Subrahmanyam (2018).

it is exogenous in our model. Further, their model is based on a single-security setting rather than a multi-security setting as in our paper and their identification thus relies on time-series data and distributional assumptions.

A key contribution to the theoretical market microstructure literature is the introduction of price-sensitive noise traders in a multi-security setting. Most models (including Kyle (1985) and Glosten and Milgrom (1985)) assume price-insensitive noise traders trading a single security. Spiegel and Subrahmanyam (1992) and Lee and Kyle (2018) use price-sensitive and risk-averse noise traders that hedge their endowment risk in a setting with one risky security (in contrast, noise trading in our model stems from liquidity-motivated and funding trading). If their noise traders were to be risk-neutral, there would be no hedging demand, and therefore no noise trading; as a result, markets would break down due to adverse selection. Kyle and Obizhaeva (2018) also model price-sensitive, risk-neutral noise traders (in essence misinformed speculators) and ensure the existence of a market equilibrium by endogenizing information production. We contribute to these few prior studies by introducing a model with risk-neutral, price-sensitive noise traders in which a market equilibrium with linear price impact can still be obtained without information production being endogenized. The main innovation to achieve this result is the multiple-security setting in which noise traders strategically optimize their trading in the cross-section in combination with investor-specific budget constraints. This innovation enables us to separate idiosyncratic order flow stemming from private information shocks from market-wide order flow stemming from liquidity shocks.

2. Model

2.1. Setup

We first introduce the basic setup for the theoretical model from which we deduce our ex post private information measure. We use underlined symbols for vectors and bold face symbols for matrices; variables without typesetting attributes are scalars. The proofs are in Appendix A. Internet Appendix IA.1 presents an overview of notation used, Internet Appendix IA.2 discusses several extensions of the basic model setup, and Internet Appendix IA.3 discusses a setting with endogenized price impact.

Our model is a one-period model with multiple investors and multiple securities. There are M investors that are strategic, risk-neutral, and indexed by i. There are N securities that are indexed by j. The prices of all securities are normalized to 1 for tractability. Each investor i is subject to a liquidity shock Z_i in dollars, where negative values refer to cash outflows and positive values to cash inflows. We assume that each investor needs to exactly accommodate this liquidity shock by trading in any (combination of) the N securities. We note that, in contrast to the myopic, price-insensitive noise traders in Kyle (1985), investors in our model are "strategic" as they optimize their uninformed trading in the cross-section of securities. For tractability, we assume that Z_i is independently and identically distributed (IID) with $E(Z_i) = 0\ \forall i$.[7]

Furthermore, each investor is subject to a vector $\underline{v}_i$ of investor-specific private information shocks for all securities j. A signal $v_{i,j}$ means that investor i receives information that security j is undervalued by $v_{i,j}$ dollars (and symmetrically overvalued by $-v_{i,j}$ dollars). The signal $v_{i,j}$ is potentially noisy. In particular, we assume that:

$$v_{i,j} = \delta_j + u_{i,j}, \tag{1}$$

where δ_j is the fundamental undervaluation of security j, and $u_{i,j}$ is a noise term with mean 0. For tractability, we assume that $E(\underline{\delta}) = \underline{0}$, such that $E(\underline{v}_i) = \underline{0}\ \forall i$. We also

[7] Independence of Z_i helps tractability as uninformed investors do not need to condition their liquidity demand on expectations about the liquidity demand by others. Relaxing the independence assumption on Z_i leaves the main results unaffected, at the expense of additional complexity.

assume that information shocks are independent of liquidity shocks.

By assumption, trading in every security j is subject to linear price impact. That is, when the total net order flow in security j equals o_j, the price impact (in dollars) equals $\lambda_j o_j$. Hence, for an individual investor, transaction costs in security j amount to:[8]

$$o_{i,j}\lambda_j o_j = o_{i,j}\lambda_j(o_{i,j}+o_{-i,j}), \tag{2}$$

where $o_{i,j}$ refers to the order flow in security j originating from investor i, $o_{-i,j}$ refers to the order flow in security j originating from all other investors, and λ_j is a security-specific price impact parameter. For the purpose of deriving optimal order flow, it is sufficient to assume λ_j to be exogenous.[9] In Internet Appendix IA.3, we validate that the assumption of linear price impact is a reasonable one in the context of our model. Specifically, we endogenize price impact in Internet Appendix IA.3, and verify that, under some conditions, an equilibrium with linear price impact indeed exists.

We make one additional assumption that is crucial for identification. We assume that there is one security for which there is no private information. Without loss of generality, we assume that this is the case for "benchmark" security $j = 1$, such that $v_{i,1} = 0\ \forall i$.

We can now write the trade optimization problem of each investor in response to the liquidity and private information shocks this investor receives as:

$$\max_{\underline{o}_i} E(\underline{\delta}|\underline{v}_i)'\underline{o}_i - \underline{o}_i'\mathbf{\Lambda}(\underline{o}_i + E(\underline{o}_{-i})), \tag{3}$$

[8] We implicitly assume that all investors use market orders for trading. While investors may in reality also use limit orders, we argue that the benefits of doing so have shrunk substantially over the past years. Kaniel and Liu (2006) derive that the benefit of using limit orders for informed trading is increasing in information horizon and decreasing with competition of other informed traders. Due to, among others, the rise of high-frequency traders, competition has increased and information horizons have shrunk.

[9] In Kyle (1985), the informed trader also takes λ as given in her trade optimization problem.

subject to the budget constraint

$$\underline{\iota}' \underline{o}_i = Z_i, \tag{4}$$

where $\underline{\iota}$ is a unity vector with all elements equal to 1, and $\mathbf{\Lambda}$ is a diagonal matrix containing all N price impact parameters. The first term in Eq. (3) represents the speculative profits due to trading on information, where negative information ($E(\delta_j|v_{i,j}) < 0$) results in selling order flow ($o_{i,j} < 0$). The second term in Eq. (3) refers to the transaction costs resulting from price impact. In order words, the investor maximizes speculative profits from private information, while minimizing transactions costs, and satisfying the budget constraint that the dollar value of all trades sums up to the liquidity shock.

We note that Eq. (3) is very similar to a classic mean-variance portfolio optimization problem. However, the quadratic disutility stems from transaction costs rather than from the variance of portfolio returns. As in classic portfolio optimization, we enforce a budget constraint specified in Eq. (4).The convenient form of the optimization problem is due to our choice of transaction costs in the form of linear price impact without fixed proportional transaction costs (such as a bid-ask spread).[10]

This basic model setup reflects a simple setting that outlines how rational investors would act in the presence of just one friction: illiquidity in the form of price impact. Since the model is simple, inverting the model yields a very tractable measure of informed trading. Our model best describes a market dominated by institutional investors. In particular, our assumptions of investor rationality, the absence of short sale constraints, and transaction costs consisting of linear price impact are relatively more applicable to

[10]If we were to (also) incorporate fixed proportional transaction costs, investors would not trade in all securities, adding substantial complexity. Our approach to focus on price impact is consistent with the notion that price impact is a relatively more important dimension of illiquidity than the bid-ask spread in today's markets.

a setting with institutional investors. This is consistent with the increasing dominance of such investors in many financial markets around the world and with the notion that these investors are more likely to possess private information (e.g., Boehmer and Kelley, 2009). The one-period horizon and the focus on market orders (as opposed to limit orders), best describe a setting in which private information is perishable, for example due to upcoming public information releases or due to learning from prices and free-riding. We are less able to capture information that is long-lived and only available to an extremely limited number of parties (as in the study of 13D filings by Collin-Dufresne and Fos, 2015), as such trades are usually more spread-out and involve more limit orders. Unreported results, however, show that our empirical results are not materially affected when we exclude the 606,127 stock-day observations within a 5-day window around 13D filings.[11]

2.2. Optimal order flow

In this subsection, we derive optimal trading strategies for each investor. For tractability, we assume for each security j that there is only a single investor with a non-zero information shock, and this information shock is perfectly accurate ($u_{i,j} = 0 \; \forall i, j$). This setup is in line with Kyle (1985). In Internet Appendix IA.2, we show that our main results also obtain when (i) there is a single informed investor, but with a noisy signal; (ii) there are multiple informed investors with a perfect signal; (iii) there are multiple informed investors with a common noisy signal; and (iv) there are multiple informed investors with a heterogeneous noisy signal. We note that in cases (ii) through (iv), $E(\underline{o}_{-i}) \neq \underline{0} \; \forall i$.

We now proceed to derive the optimal trading behavior of individual investors exposed to liquidity and private information shocks. In Section 2.2.1, we consider the case of liquidity shocks only. In Section 2.2.2, we add private information shocks.

[11] We are grateful to Pierre Collin-Dufresne and Vyacheslav Fos for providing their data on 13D filings.

2.2.1. Liquidity shock only

In this subsection, we impose that the private information shocks in all securities are equal to 0 for all investors and investor i's only demand for trading stems from a liquidity shock (that is, $\underline{v}_i = \underline{0}$ and $Z_i \neq 0$). Since the liquidity shock needs to be accommodated in full, the investor incurs transaction costs. Intuitively, the investor minimizes transaction costs when the marginal utility (marginal cost) of trading is equal across all securities. When this first-order condition holds, it is impossible for order flow to be re-allocated across securities to reduce total transaction costs. Mathematically, we solve the constrained optimization problem in Eq. (3) with a Lagrangian Multiplier technique to obtain:

Lemma 1. *Optimal order flow for investor i is given by:*

$$\underline{o}_i = Z_i \mathbf{\Lambda}^{-1} \underline{\iota} (\underline{\iota}' \mathbf{\Lambda}^{-1} \underline{\iota})^{-1}. \tag{5}$$

Proof. See Appendix A. $\square$

The solution derived in Lemma 1 has the following intuition. The marginal cost of trading security j equals $2\lambda_j o_{i,j}$. Hence, to have equal marginal trading costs across all securities, investor i needs to set order flow for security j inversely proportional to the price impact parameter λ_j. This needs to be scaled to make sure that the total order flow for investor i adds up to liquidity shock Z_i. The appropriate scaling parameter is the constant $(\underline{\iota}' | \mathbf{\Lambda}^{-1} \underline{\iota})$, which equals the sum of all inverse price impact parameters. The expression $\lambda_j / (\underline{\iota}' | \mathbf{\Lambda}^{-1} \underline{\iota}) \in (0, 1)$ then represents the fraction of the liquidity Z_i that is accommodated in security j.

There are several other things to be noted about the optimal order flow in Eq. (5). First, investor-specific parameters only show up through the liquidity shock Z_i, which is a scalar. Hence, following a liquidity shock, the proportions in which the different securities are traded are identical across investors and scale with the magnitude of their liquidity shocks (in dollars). Second, the amount of liquidity-motivated trading in a security is inversely proportional to price impact, such that investors optimally spread their trading across all securities and the most liquid securities see most trading.

2.2.2. Adding an information shock

In this subsection, each investor i receives – in addition to a liquidity shock Z_i – an investor-specific private information shock $\underline{v}_i$, with $v_{i,j} \neq 0$ for at least one security j. Since the information is perfectly accurate in the baseline model by assumption, $E(\underline{\delta}|\underline{v}_i) = \underline{v}_i$. Intuitively, as in Section 2.2.1, solving the constrained optimization problem in Eq. (3) requires the marginal utility of trading each security to be equal. The marginal utility of trading each security j is now given by $v_{i,j} - 2\lambda_j o_{i,j}$. As a result, the order flow in security j will be tilted in the direction of the signal – even if the price pressure stemming from the liquidity shock in security j works in the same direction and even if such price pressure is large. Mathematically, we again use the Lagrangian Multiplier technique to obtain the optimal order flow for investor i:

Lemma 2. *Optimal order flow for investor i is given by*

$$\underline{o}_i = Z_i \mathbf{\Lambda}^{-1} \underline{\iota} (\underline{\iota}' \mathbf{\Lambda}^{-1} \underline{\iota})^{-1} + \frac{1}{2} \mathbf{\Lambda}^{-1} (\mathbf{I} - (\underline{\iota}' \mathbf{\Lambda}^{-1} \underline{\iota})^{-1} \underline{\iota}\underline{\iota}' \mathbf{\Lambda}^{-1}) \underline{v}_i. \tag{6}$$

Proof. See Appendix A. □

We can understand the intuition of Eq. (6) as follows. The first term in this equation is identical to Eq. (5), which specifies how the liquidity shock Z_i is allocated across securities. The second term represents the "speculative order flow" in each security j due to private information about that security. Since transaction costs are quadratic, this speculative demand is inversely proportional to the security-specific price impact parameter λ_j. This basic result on the intensity of speculative trading is also present in the single-security setting of Kyle (1985, see his Eq. (2.6)).

The third term is the "funding order flow" that is required (because of the budget constraint) to finance the speculative order flow. This funding demand results in additional uninformed order flow. Specifically, investor i trades all other securities in the opposite direction to finance the speculative trade in security j. This term is most conveniently written as $(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{v}_i)\mathbf{\Lambda}^{-1}\underline{\iota}(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}$. In this form, the funding need is equivalent to an additional liquidity shock to investor i. The the size of the funding need equals the sum of speculative demands for investor i across all securities j and is given by $\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{v}_i$, which is a scalar. As in Section 2.2.1, this amount is multiplied by a vector that indicates how investors optimally spread their funding trading across all securities: $\mathbf{\Lambda}^{-1}\underline{\iota}(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}$.

2.3. Reverse-engineering private information

In this subsection, we reverse-engineer the private information signals for each security j from the aggregate order flow of all investors i, which isthe ultimate goal of our model. To this end, we first pre-multiply both sides of Eq. (6) by $\mathbf{\Lambda}$ and re-arranging terms:

$$\begin{aligned}\mathbf{\Lambda}\underline{o}_i &= \frac{1}{2}\underline{v}_i+(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}(Z_i\underline{\iota}-\frac{1}{2}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{v}_i),\\ &= \frac{1}{2}\underline{v}_i+\underbrace{(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}}_{\text{scalar}}\ \underbrace{(Z_i-\frac{1}{2}\sum_j \lambda_j^{-1}v_{i,j})}_{\text{scalar}}\,\underline{\iota},\end{aligned} \tag{7}$$

where we use the fact that $\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{v}_i = \sum_j \lambda_j^{-1}v_{i,j}$, which represents the aggregate funding demand of investor i generated by speculative trading in all securities j and is a scalar.

Eq. (7) is an important result, as it shows that ranking securities on private information shocks is equivalent to ranking them on the product of their price impact parameter and order flow, and that this is true for each investor i.

The second term on the right hand side of Eq. (7) refers to uninformed order flow components (liquidity-motivated and funding order flow), and can be expressed as a unit vector multiplied with scalars, implying that it is identical for all securities. Since we assume that benchmark security $j = 1$ is insulated from private information shocks (that is, $v_{i,1} = 0 \ \forall i$), the product of the two scalars in the second term on the right-hand side can be exactly identified by $\lambda_1 o_{i,1}$, obtained through substituting $v_{i,1} = 0$ in the first row of Eq. (7). As a result, we can derive a "correction term" for all uninformed (liquidity-motivated and funding) trading by investor i that can be subtracted from all other rows of Eq. (7) to "purge" all rows of uninformed order flow. We note that this is true for all investors i. We formalize this result in the following Lemma:

Lemma 3. *The vector of investor-specific private information shocks can be extracted from the order flow of investor i in all securities j by:*

$$\underline{v}_i = 2\mathbf{\Lambda}\underline{o}_i - 2\lambda_1 o_{i,1}\underline{\iota}. \tag{8}$$

Moreover, $\mathbf{\Lambda}\underline{o}_i$ (without the correction term that is equal for all securities j) preserves the cross-sectional rank ordering of $\underline{v}_i$.

Proof. See Appendix A. □

Eq. (8) is a powerful result. Uninformed trading always takes place in all securities in fixed proportions. Since there is never any informed trading in security 1 (by assumption), the order flow in security 1 multiplied by its price impact parameter λ_1 is a sufficient statistic for the uninformed trading in *all securities*. We can use this sufficient statistic to purge the order flow in each security j of its uninformed component and thus isolate its informed component. Since informed order flow is inversely proportional to the price impact parameter λ of security j, we can subsequently use λ_j to reverse-engineer the private information shock about security j received by investor i ($v_{i,j}$) from the informed component of investor i's order flow in security j. There is also a revealed preference interpretation to this result. Investors only trade a very illiquid security when it is very profitable to do so. Hence, sizable order flow in an illiquid security is an indication of valuable private information.

Summarizing, if an information-free security exists, we can extract investor-specific information shocks using security-specific price impact parameters and investor-specific order flow. For cross-sectional analyses, the product of price impact and investor-specific order flow suffices to rank securities on their investor-specific private information shocks.[12]

Although the result in Eq. (8) provides valuable insights, it does not yield a broadly implementable measure of private information, since we typically do not observe order flow at the investor-level. However, the price impact parameters in Eq. (8) are security-specific but not investor-specific. As a result, Eq. (8) is linear in order flow with coefficients that are not investor-specific, which allows for easy aggregation:

Proposition 1. *The vector of private information shocks can be extracted from the aggregate order flow across all investors i in all securities j by:*

$$\underline{v} = 2\mathbf{\Lambda}\underline{o} - 2\lambda_1 o_1 \underline{\iota}, \tag{9}$$

where $\underline{o} = \sum_i \underline{o}_i$, $o_1 = \sum_i o_{i,1}$, and $\underline{v} = \sum_i \underline{v}_i$.

Proof. See Appendix A. □

Eq. (9) is the market-level equivalent of Eq. (8) and only contains elements that are observable (or can be estimated) at the market level. This equation is the basis for our

[12] We note that if no information-free security exists, the system of equations in Eq. (6) is underidentified and private information shocks cannot be exactly identified. However, in that case, the rank-ordering of investor-specific private information shocks is still preserved by the rank ordering of $\mathbf{\Lambda}\underline{o}_i$.

private information measure. In particular, Eq. (9) says that we can estimate the private information shock in security j from the price impact parameters and aggregate order flows in all securities (aggregated across all investors) using the following expression: $2\lambda_j o_j$–$2\lambda_1 o_1$ (where the cross-sectional rank ordering of the private information shock is preserved when using the simplified expression $\lambda_j o_j$; that is, without the correction term for aggregate uninformed trading $2\lambda_1 o_1$ and without the multiplicative factor 2).

In Internet Appendix IA.2.1 through Internet Appendix IA.2.4, we show that similar results obtain, up to a scalar multiplication, in the presence of noise in signals and with multiple informed investors. In particular, we show that as the number of investors with private information grows from one in our baseline model to k in the model extensions, our measure changes to $\underline{\bar{v}} = \frac{k+1}{k}(\mathbf{\Lambda}\underline{o} - \lambda_1 o_1 \underline{\iota})$, such that $\lim_{k\to\infty} \underline{\bar{v}} = \mathbf{\Lambda}\underline{o} - \lambda_1 o_1 \underline{\iota}$, where $\underline{\bar{v}}$ is the average private information shock across informed investors. Hence, the measure from the baseline model ($2\lambda_j o_j$–$2\lambda_1 o_1$) represents one (k=1) to two ($k \to \infty$) times the underlying private information shock. Internet Appendix IA.2.5 provides a further robustness test of the model. In particular, we show that the main result also obtains when a subset of investors has a transaction cost-free bank account which can be used to accommodate liquidity shocks and funding needs for speculative trades.

In this section, we have shown how our trade optimization yields a simple and intuitive expression for a security's private information shock; in cross-sectional applications this expression reduces to $\lambda \times o$. We note that the microstructure literature offers various ways to estimate a security's price impact parameter λ and that a security's aggregate order flow o from the model can be measured empirically by its order imbalance (OIB), which implies that our private information measure can be used empirically in a straightforward way. In the remainder of the paper, we set out to estimate and validate our measure for a large sample of U.S. stocks over a prolonged time period.

3. Data and variable definitions

For our empirical analysis of the model introduced in Section 2, we use all NYSE-listed common stocks – identified as CRSP PERMNOs with share codes 10 or 11 – that survive our data filters (as described below). We include only NYSE stocks to prevent issues with differences in trading mechanisms across NYSE and Nasdaq (Gao and Ritter, 2010). Our sample starts on February 1, 2001 (to prevent issues stemming from the tick size change on January 29, 2001) and runs until the end of 2014.

3.1. Estimates of order imbalance and price impact

We run all our analyses at the daily frequency and thus estimate the key variables needed for our private information measure (order imbalance and price impact) for each stock each day based on intraday data. We obtain national best bid and offer prices (NBBO) and transaction data across all U.S. exchanges for individual stocks from Thomson Reuters Tick History (TRTH).[13] We discard stocks that trade for less than \$5 at any time during our sample period and filter all quote and trade data as in Rösch, Subrahmanyam, and van Dijk (2017).

We determine the sign of each trade using the Lee and Ready (1991) algorithm, as follows. If a trade is executed at a price above (below) the quote mid-point, we classify it as a buy (sell). If a trade occurs exactly at the quote mid-point, we sign it using the previous trade price according to the tick test. That is, we classify the trade as a buy (sell) if the sign of the last price change is positive (negative). If the price is the same

[13]TRTH is a record of Thomson Reuters' worldwide real-time Integrated Data Network (IDN), and prices are time-stamped with at least millisecond precision. Because of the high-precision time-stamps we do not need to adopt the suggested algorithm by Holden and Jacobsen (2014) to deal with second-by-second timestamps as in the monthly TAQ data. The TRTH database is used in several recent studies (e.g., Fong, Holden, and Trzcinka (2017); Kahraman and Tookes (2017); Lau, Ng, and Zhang (2012); Lai, Ng, and Zhang (2014); Marshall, Nguyen, and Visaltanachoti (2011)). Rösch, Subrahmanyam, and van Dijk (2017, footnote 6) show that differences between TRTH and TAQ are very small for NYSE stocks.

as the previous trade (a zero tick), then the trade is a zero-uptick (zero-downtick) if the previous price change was positive (negative), and thus classified as a buy (sell). If the previous price change was also equal to zero, we discard the trade. Following recent papers, we do not use a delay between a trade and its associated quote because of the decline in reporting delays (see Madhavan, Richardson, and Roomans, 1997; Chordia, Roll, and Subrahmanyam, 2005). We are able to sign the overwhelming majority of trades in this way. For each stock on each day, we then compute its order imbalance (OIB) as the dollar volume of buyer- minus seller-initiated trades based on the signed trades over that day. We express order imbalance in millions of US$.

We estimate the daily price impact parameter for each stock following Goyenko et al. (2009) and Hasbrouck (2009). For each stock on each day, we estimate a regression of five-minute log-returns from mid-quote prices on signed dollar trading volume in the same five-minute interval. The slope coefficient from this regression represents the variable costs of trading and can be interpreted as the stock's price impact parameter, in the spirit of Kyle's (1985) lambda. In contrast to Goyenko et al. (2009, their Eq. (5)), we do not take the square-root of the dollar trading volume in this regression, such that our price impact measure has a straightforward interpretation: the percentage price change per unit of dollar trading volume.[14]

We discard stock-days with fewer than 50 trades to ensure a minimum number of observations to estimate this price impact regression. Nonetheless, individual price impact estimates are noisy and could lead to extreme estimates in our measures of informed trading and private information. Furthermore, our model assumes that investors optimize securities trades following liquidity and private information shocks based on the securi-

[14]Note that, in the model, prices are normalized to one and hence, price impact in terms of returns and dollars coincide. Accommodating for heterogeneous price levels across stocks leads to price impact in terms of returns, similar to how the Amihud (2002) illiquidity measure is defined.

ties' *expected* price impact. In other words, estimating price impact parameters over the same day as we measure order imbalances (which, within the model, arise as a result of the trading by individual investors) might introduce look-ahead bias. To mitigate these concerns, we construct measures of the expected price impact of trading a given stock on a given day (λ) as the moving average of the estimated daily price impact parameters for that stock over the past 20 trading days, where we set negative price impact estimates to zero. To further reduce the influence of outliers, we cross-sectionally winsorize the resulting expected price impact estimates each day at the 95% level.

3.2. Estimates of private information measures

We estimate the private information shock – or v – for each stock on each day based on Eq. (9) as twice the $\lambda\times OIB$ of the stock of interest minus the $\lambda\times OIB$ of the benchmark security. This measure of the private information shock underlying informed trading is thus based on just the estimates of the λ and OIB of the stock of interest and of the benchmark security. In cross-sectional applications, the benchmark's term drops out since it is common to all individual securities over a given time period. Furthermore, dropping the multiplication by 2 preserves the cross-sectional rank ordering, thus yielding a very simple measure of private information: $\lambda\times OIB$. Both v and $\lambda\times OIB$ not only indicate the magnitude of the private information shock, but also its sign – and can thus assume both positive and negative values.

Throughout our empirical analyses, we distinguish between our private information measure and PIN (Easley et al., 2002). We estimate traditional PIN using a three-month rolling window using the R package "pinbasic". PIN is commonly estimated at lower frequencies than the daily frequency used in this paper and is also not signed. To put it on equal footing with our daily (signed) $\lambda\times OIB$ measure, we follow Brennan et al. (2018) and estimate their daily (signed) posterior probabilities of informed buying and selling (labeled *Good PIN* and *Bad PIN*, respectively) per stock-day using the number of buy and sell trades on the current day and the previously estimated traditional PIN measure as a prior. In unreported robustness tests, we control for $VPIN$ (developed by Easley et al., 2012) instead of PIN, and obtain similar results.

3.3. Estimates of other variables, final sample, and benchmark security

In our empirical analyses, we also use daily mid-quote returns computed from the daily mid-point of the last quote on each day, adjusted for corporate actions using CRSP data, and cross-sectionally winsorized each day at the 0.1% and 99.9% level (*Return*). For some of our tests, we use a spread-based liquidity measure computed as the difference between the quoted ask and the quoted bid price scaled by the mid-point of the quotes, averaging the spread associated with all trades for the stock on that day (proportional quoted spread or $PQSPR$). We also compute the market capitalization (*Mktcap*) of each stock based on the number of shares outstanding and prices from CRSP at the beginning of each calendar year. After constructing these variables, we drop stocks with fewer than six months of data over our sample period. In addition, when the data for a stock exhibit a gap of more than two months, we only retain the longest uninterrupted period. Our final sample consists of all 1,388 NYSE stocks that survive these data screens over 2001-2014 and is based on a total of 18,626,168,999 signed trades.

We use the SPDR S&P500 ETF (ticker "SPY") as a benchmark security that is insulated from informed trading. Our motivations for choosing the SPDR as the benchmark are that it is highly traded and that it seems unlikely that informed traders exploit their private information by trading such a passive market-wide benchmark.[15] However, since most of our analyses are cross-sectional in nature, we can use the simplified $\lambda \times OIB$ mea-

[15]This argument is similar to the rationale behind program trading facilities. Such facilities allow better liquidity because multiple securities need to be traded at the same time and hence the likelihood of trading on private information about any of these securities is low.

sure that does not include the benchmark's term. Consolidated trades and quotes for the SPDR are from the TRTH database. We estimate the order imbalance and the price impact parameter of the benchmark security in the same way as we do for individual stocks, based on 783,175,721 signed trades for the SPDR.

3.4. Summary statistics

Table 1 presents summary statistics of the daily *Return*, $PQSPR$, λ, OIB, v, $\lambda \times OIB$, *Good PIN*, and *Bad PIN* across all stocks in our sample over 2001-2014. The table reports cross-sectional summary statistics (mean, standard deviation, median, and 25th and 75th percentiles) of the stock-by-stock time-series averages of these variables. The table is based on all 1,388 NYSE stocks in the sample, for which we have daily observations for 1,992 days on average. We note that the more stringent data requirements to estimate PIN lead to a somewhat lower number of stocks (1,260) for the *Good PIN* and *Bad PIN* measures.

The mean and median mid-quote returns are equal to, respectively, six and five basis points per day. The median $PQSPR$ is 11 basis points. The median λ (which is scaled by 10^6 to express price impact as the percentage price change associated with a million dollars of trading volume) equals 0.51%, which means that the median of the average price impact across all stocks in the sample is 51 basis points for a trade of \$1m. The median OIB is slightly positive (\$0.35m.) over our sample, and exhibits substantial cross-sectional variation, with a standard deviation of \$2.41m. The mean OIB and λ of the SPDR benchmark security are equal to, respectively, \$30.29m. and 0.21 basis points per \$1m trade per 5-minutes (not tabulated). These numbers indicate that the SPDR experienced substantial inflows over our sample period and that the average price impact of trading the SPDR is tiny, at less than one 100th of the cross-sectional mean of the average price impact of all NYSE stocks in our sample of 0.83%.

The mean and median v are equal to 0.30 and 0.19, respectively, which suggests that the private information shock was slightly positive in our sample. The cross-sectional standard deviation of the mean v across stocks is substantial, at 0.62. The mean and median $\lambda\times OIB$ (if multiplied by two as in v) are slightly higher than the mean and median v, indicating that the correction term for aggregate uninformed trading is positive over our sample period. The mean of *Good PIN* and *Bad PIN* are equal to 19% and 14%, respectively, comparable in magnitude to the mean PIN estimate of around 19% reported by Easley et al. (2002).

To get a sense of the time-series variation in private information in our sample, we plot the average v of the top and bottom decile portfolios of stocks sorted on v each day in Figure 1. In the early years of our sample period, the private information shock is relatively pronounced and tends to be somewhat larger in magnitude for stocks with positive private information shocks than for stocks with negative private information shocks. The average magnitude of the private information shocks decreases slowly over time in 2003-2007 (both for positive and negative shocks), after which it shows a peak in the period surrounding the start of the financial crisis in 2008-2009 (especially for negative shocks), to return to pre-crisis levels by 2010.

Table 2 shows the pooled contemporaneous correlations between the private information measure v, the absolute value of this measure $|v|$, the simplified private information measure $\lambda\times OIB$, the absolute simplified private information measure, $\lambda\times|OIB|$[16], λ, OIB, $PQSPR$, trading volume, *Return*, *Good PIN* and *Bad PIN* (with p-values in parentheses). v is very highly correlated with its simplified version $\lambda\times OIB$, at 0.935 (and the correlation between the absolute values of these measures is similar, at 0.932). Our model prescribes that, when comparing private information over time, the $\lambda\times OIB$

[16]Because λ is non-negative by construction, we only need to take the absolute value of OIB.

of the benchmark security needs to be subtracted from the $\lambda \times OIB$ of the security of interest. However, the very high correlation between v and $\lambda \times OIB$ suggests that, in practice, the simplified version might also suffice in time-series analyses.[17]

Table 2 also shows that λ has a slight negative correlation with $\lambda \times OIB$ (at -0.031), and that the correlation between OIB and $\lambda \times OIB$ is also far from perfect (at 0.489), which suggests that our simplified private information measure is distinct from its individual components and that any results we find for $\lambda \times OIB$ are unlikely to stem solely from λ or OIB. As expected, a stock's quoted spread is positively correlated to the absolute magnitude of private information in that stock as well as to the stock's price impact parameter. OIB is negatively correlated with both $PQSPR$ and λ. The correlations with returns provide some initial evidence that our measures pick up meaningful variation in private information, since both v and $\lambda \times OIB$ are positively and significantly related to contemporaneous stock returns.

Furthermore, it is comforting to see that the correlations of our private information measures v and $\lambda \times OIB$ with *Good PIN* and *Bad PIN* are in line with expectations (positive with *Good PIN* and negative with *Bad PIN*) and statistically significant, although the small magnitudes of these correlations indicate that the overlap of our new measures with PIN is limited.

4. Empirical results

In this section, we present and discuss the results of six empirical tests to examine whether the private information measure stemming from our model can be applied to real-life data and yields results that are consistent with our theoretical interpretation of this measure. We examine the characteristics of stocks with large vs. small private

[17] In unreported tests, we find that the very high pooled correlation between v and $\lambda \times OIB$ reported in Table 2 is not only due to the cross-section. In particular, the within-R^2 of a panel model with stock fixed effects of v on $\lambda \times OIB$ is 0.87, corresponding to a time-series correlation of around 0.9.

information shocks (Section 4.1), the relation between the private information measure and contemporaneous stock returns (Section 4.2), return reversals (Section 4.3), and future return volatility (Section 4.4) as well as the behavior of the private information measure around M&A announcements (Section 4.5) and the exogenous loss of analyst coverage (Section 4.6).

4.1. Characteristics of stocks with large vs. small private information shocks

As a first indication of the empirical attributes of our private information measure, we examine the characteristics of stocks with large vs. small private information shocks. We sort stocks into quintile portfolios each month based on the average value of their absolute private information shock $|v|$ across the days within that month and compare the characteristics of stocks in these quintiles. We expect that smaller stocks with greater uncertainty exhibit a greater degree of private information, and thus larger absolute values of our private information measure.

Table 3 presents the results of these quintile sorts and shows the average size (market capitalization in \$bln), return volatility, bid-ask spread ($PQSPR$), price impact (λ), analyst coverage (number of analysts following the stock from I/B/E/S), and analyst dispersion (standard deviation of annual analysts' earnings per share forecasts scaled by the mean forecast from I/B/E/S) for each of the five quintile portfolios sorted on $|v|$ as well as the difference in these characteristics between Quintile 5 (large absolute private information measure) and Quintile 1 (small absolute private information measure).[18]

[18]In Table 3, we sort on $|v|$ instead of $\lambda \times |OIB|$ since here we first average the private information measure across days within the month and thus the benchmark term does not exactly drop out in cross-sectional comparisons. Since the remainder of the paper focuses on cross-sectional analyses that do not suffer from this minor issue, we use the simplified private information measure $\lambda \times OIB$. To prevent that our results in Table 3 are driven by private information potentially affecting the stock characteristics, these characteristics are taken from the month prior to the quintile sort based on $|v|$. However, these characteristics are all very persistent and we obtain similar results when we take them from the month of the sort.

In line with expectations, we find that stocks with greater absolute values for the private information measure tend to have a smaller market capitalization, greater return volatility, and a larger bid-ask spread. The differences in market cap and bid-ask spread across Quintiles 5 and 1 are relatively modest, but the difference in volatility is large: stocks in Quintile 5 are more than twice as volatile as stocks in Quintile 1 (9.15% vs. 4.09%). Stocks with larger absolute private information estimates also have considerably larger price impact parameters. This result may not be surprising, since λ enters our private information measure directly. At the same time, the pooled correlation between λ and $\lambda \times |OIB|$ presented in Table 2 is small, and the findings in Table 3 are consistent with the notion that stocks with larger private information shocks exhibit a greater degree of asymmetric information as reflected in their price impact.

Somewhat surprisingly, stocks in Quintile 5 tend to have a slightly higher number of analysts than stocks in Quintile 1, but the difference is small (11.8 vs. 11.0 analysts on average). We note that the NYSE stocks in our sample are relatively large and most have substantial analyst following, and thus the variation in analyst following within our sample is relatively limited. Also, there may be a greater need for analysts when asymmetric information is more severe.

Perhaps more interestingly, there is a large difference in the analyst dispersion across stocks in Quintiles 5 and 1. Stocks with large absolute information shocks tend to have analyst dispersion that is three times as large as stocks with small absolute information shocks (0.15 vs. 0.05), suggesting considerable more scope for information asymmetries for those stocks.

In sum, we interpret the results in Table 3 as consistent with the notion that the degree of private information as captured by our measure is larger for smaller stocks with more uncertainty for which there is more scope for information asymmetries.

4.2. Relation of private information shocks with contemporaneous returns

We now turn to a straightforward but important premise of our private information measure. If our measure truly picks up the private information shock underlying informed trading in individual stocks, we would expect a strong positive contemporaneous relation between the private information measure and stock returns. After all, informed trading should lead stock prices to at least partially incorporate the underlying private information shock on the same day.

Figure 2 provides a first indication of the relation between our private information measure and contemporaneous stock returns by plotting the time-series of the returns of the top and bottom decile portfolios of stocks sorted on v each day (from Figure 1). The patterns in Figure 2 are a near mirror image of those in Figure 1, indicating that the contemporaneous returns of stocks with positive (negative) private information shocks tend to be positive (negative) and that the strength of this relation is relatively stable over time. The economic magnitude of this relation is substantial: the top decile portfolio of stocks sorted on v each day has an average return of around 1% per day over our sample period, while the bottom decile has an average return of -1% per day. We note that, as our measure is constructed ex post based on realized OIB on the same day, a trading strategy exploiting these return differentials is not implementable.

In Table 4, we substantiate the initial evidence from Figure 2 on the positive association between our private information measure and contemporaneous returns by running daily Fama and MacBeth (1973) regressions of the daily mid-quote returns of individual stocks on one-day lagged returns, our contemporaneous simplified private information measure $\lambda\times OIB$, as well as the individual components of this measure λ and OIB.

Consistent with prior studies, we find that daily stock returns exhibit significantly negative autocorrelation (e.g., Roll, 1984; Cox and Peterson, 1994; Nagel, 2012). The coefficient on lagged returns is equal to -0.020 in the first model in Table 4, with a Fama-MacBeth t-stat of -9.21 (based on the Newey and West, 1987, correction for autocorrelation in the estimated coefficients).

More importantly, we find a positive and highly significant effect of our simplified private information measure $\lambda \times OIB$ on contemporaneous stock returns. This result suggests that returns are significantly higher (lower) for stocks with a more positive (negative) value of $\lambda \times OIB$ on that day, which is what we would expect if $\lambda \times OIB$ measures private information. The second model of Table 4 shows that the effect of $\lambda \times OIB$ is not driven by λ or OIB itself, and that its t-stat is considerably higher than the individual t-stats of the coefficients on λ or OIB. In other words, our new private information measure is more than the sum of its two well-known parts.

The economic magnitude of this effect is considerable. A one standard deviation increase in $\lambda \times OIB$ is associated with a 0.19 standard deviation increase in contemporaneous stock returns (based on the second model in Table 4), which is substantial in light of the noise inherent in daily stock returns.

In the third and fourth regression models of Table 4, we examine whether the effect of $\lambda \times OIB$ disappears when we introduce other "scaled" versions of order imbalance that may be correlated with $\lambda \times OIB$. In the third model, we include the product of OIB and the inverse of a stock's market capitalization, to reflect the idea that order imbalance may move prices more for small stocks. In the fourth model, we include the product of OIB and $PQSPR$ to examine whether other liquidity proxies subsume our main finding. Although the coefficients of both $OIB \times 1/Mktcap$ and $OIB \times PQSPR$ are positive and significant, the effect of $\lambda \times OIB$ remains intact. Further, we examine whether the contemporaneous relation between stock returns and our simplified private information measure is robust to including *Good PIN* and *Bad PIN* in the final two models of Table 4 (excluding and including the other scaled versions of order imbalance, respectively). Although the coefficients on *Good PIN* and *Bad PIN* are significant with the expected sign, the coefficient of $\lambda \times OIB$ is hardly affected and still strongly significant.

Overall, the results in Table 4 indicate that $\lambda \times OIB$ has explanatory power for the cross-section of returns that goes beyond that of λ and OIB individually. We are not aware of theoretical models that provide an alternative interpretation of $\lambda \times OIB$ than the interpretation as a proxy for private information that our model suggests. Furthermore, the explanatory power of $\lambda \times OIB$ is not subsumed by other "scaled" measures of order imbalance and/or by PIN.

4.3. Relation of private information shocks with return reversals

Although the evidence in Table 4 is consistent with our $\lambda \times OIB$ measure capturing private information, it is not sufficient to validate our measure since it is based on the contemporaneous relation of $\lambda \times OIB$ with returns that could in part be driven by liquidity-motivated trading, which may also move prices. Therefore, we now turn to potentially more stringent tests of our conjecture that $\lambda \times OIB$ measures private information that are more powerful in discriminating between liquidity-motivated and informed trading.

In particular, a crucial distinction between these two types of trading is that informed trading should be associated with permanent price impact, while the price impact of liquidity-motivated trading should be transitory and should thus result in subsequent return reversals (e.g., Kyle, 1985; Admati and Pfleiderer, 1988; Glosten and Harris, 1988). Therefore, we would expect to observe significantly weaker return reversals following stock-days with larger positive or negative values of $\lambda \times OIB$.

To test this hypothesis, Table 5 reports the results of daily Fama-MacBeth regressions of the daily mid-quote returns of individual stocks on one-day lagged returns, as well as one-day lagged returns interacted with one-day lagged $\lambda \times |OIB|$. If returns revert significantly less following information shocks, and if our measure is a meaningful proxy for these shocks, the coefficient on the interaction term should be positive. We note that we take the absolute value of our simplified private information measure $\lambda \times OIB$ for these tests, since return reversals should be weaker following larger positive *or* negative information shocks. Consistent with Table 4, the first-order autoregressive coefficient is significantly negative, at -0.008 in the first model of Table 5. In the second model, we add lagged $\lambda \times |OIB|$ as well as lagged $\lambda \times |OIB|$ interacted with lagged returns. The coefficient on lagged $\lambda \times |OIB|$ is positive and significant, suggesting that returns tend to be higher for stocks with a more pronounced previous-day private information shock.[19]

The coefficient on the key interaction term of lagged returns and lagged $\lambda \times |OIB|$ in the second model of Table 5 is significantly positive at 33.547, with a Fama-MacBeth Newey-West t-stat of 3.75. This finding indicates that, indeed, stock returns tend to revert significantly less following stock-days with higher absolute values of our private information measure. One question that may arise is how much our proxy for private information shocks adds to using λ and OIB individually. To address this question, we again break up $\lambda \times |OIB|$ into its two individual components. The third model of Table 5 shows our reversal analysis using only λ and OIB separately. In the fourth model, we then add back $\lambda \times |OIB|$ and its interaction with lagged returns. We find that adding $\lambda \times |OIB|$ increases the adjusted R^2 by around 20%, from 3.21% in the third model to 3.85% in the fourth model. Moreover, the coefficient on the interaction term of lagged returns with λ has the "wrong" sign in both the third and the fourth model, suggesting that $\lambda \times |OIB|$ cannot be replaced by λ alone as an ex post private information measure.

Taken together, we interpret this evidence as consistent with the view that $\lambda \times OIB$ proxies for private information shocks, does not capture the price impact of liquidity-

[19]This effect may be driven by our finding in Figure 1 that, over our sample period, positive information shocks tend to be somewhat greater than negative shocks. An alternative interpretation is that stocks that are more subject to informed trading command higher expected returns.

motivated trading, and contains non-negligible additional information relative to its components λ and OIB. In the fourth and fifth models of Table 5, we add the one-day lagged maximum of *Good PIN* and *Bad PIN* (as the equivalent of "absolute" private information) as well as its interaction with lagged returns – excluding and including the interaction terms stemming from the two components of $\lambda \times |OIB|$. We find weak evidence that this additional interaction term has a significant coefficient (albeit with unexpected sign), but our main result that return reversals are weaker following stock-days with greater absolute $\lambda \times OIB$ is not affected.

To assess the economic significance of the reduced strength of return reversals following large absolute private information shocks, we also take a double-sorting approach to studying the relation between the private information measure and return reversals. We first sort stocks into quintile portfolios on day $d-1$ based on their $\lambda \times |OIB|$. Quintiles 1 and 5 thus contain stocks with, respectively, small and large absolute private information estimates on that day. Subsequently, we sort stocks within each private information quintile into five subportfolios based on their returns on day $d-1$. We then compute the returns on a simple reversal strategy within each absolute private information quintile that is long in day $d-1$'s loser stocks (subportfolio 1) and short in day $d-1$'s winner stocks (subportfolio 5) in that quintile. The returns of the reversal strategy are based on these stocks' next day's returns computed from the market close on day $d-1$ till the market close on day d. Comparing the abnormal returns on the reversal strategies within Quintile 1 and within Quintile 5 allows us to assess the economic significance of the difference in the strength of return reversals between stocks with small and large absolute private information shocks.

The results are in Table 6. The first four columns report the estimates of time-series regressions of the daily returns on the reversal strategy for private information Quintile 1 (small absolute private information shocks) on various asset pricing factors. The columns correspond to, respectively, the CAPM, the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, and the Carhart model supplemented with a fifth factor based on short-term reversals (Jegadeesh, 1990). Daily returns on these factors are from the website of Ken French. All four models indicate economically large and statistically highly significant abnormal returns (alphas) for Quintile 1 of 7-8 basis points per day (around 18% per annum) with t-stats greater than 4, which indicate strong daily return reversals for stocks for which we observe little or no private information.[20]

The fifth column of Table 6 shows the results of a similar regression but then for Quintile 5 (large absolute private information shocks). For brevity, we report results for the five-factor model only. The five-factor alpha of the reversal strategy within Quintile 5 is small and statistically indistinguishable from zero (alpha of 0.7 basis point per day, with a t-stat of 0.34). This finding indicates that returns do not revert following stock-days with large absolute private information shocks, and thus that the price impact of informed trading (as documented in Section 5.2) is permanent rather than transitory. The final column shows the five-factor alpha of the difference in the reversal strategy returns between Quintiles 1 and 5. This column shows that, indeed, the difference in the strength of return reversals between stocks with small and large absolute private information shocks is economically substantial and statistically significant (alpha of 6.3 basis points per day, with a t-stat of 3.59).

Overall, our tests in Section 5.3 show that stocks with very negative or very positive private information estimates subsequently exhibit significantly weaker return reversals than stocks with small private information estimates, supporting the theoretical interpretation of our new private information measure.

[20]These abnormal return estimates on reversal strategies are somewhat lower than the mean reversal returns reported by Nagel (2012) of 18 basis points per day (also based on mid-quote returns), likely because our sample comprises of larger stocks as we only include NYSE stocks and impose relatively strict data screens for the intraday data.

4.4. Relation of private information shocks with future return volatility

Our fourth empirical test evaluates the hypothesis that informed trading should intensify ahead of periods of high return volatility. Johnson and So (2018) outline three motivations for this hypothesis derived from various theoretical models of informed trading (e.g., Grossman and Stiglitz, 1980; Easley and O'Hara, 1987; Easley et al., 2012). First, if informed traders possess private information about upcoming news, they may trade ahead of the news and return volatility may increase as the news becomes public. Second, for a given number of informed traders, higher volatility news may imply a more profitable trading opportunity and thus greater informed order flow. Third, informed traders may create an order imbalance cascade that results in higher return volatility. All three motivations predict that private information as captured by our measure should predict future stock return volatility.

In line with Johnson and So (2018), in Table 7 we present the results of daily Fama-MacBeth regressions of the daily realized return volatility (squared return) of individual stocks on one-day lagged returns, five daily lags of realized return volatility, one-day lagged absolute $\lambda \times |OIB|$, as well as its components one-day lagged λ and one-day lagged $|OIB|$. The results indicate that absolute $\lambda \times |OIB|$ positively and significantly predicts realized return volatility, even after controlling for lagged returns, lagged volatility, and lagged λ and lagged $|OIB|$ separately.

Similar to Table 4, models (4) and (5) of Table 7 include other "scaled" versions of absolute order imbalance that may be correlated with absolute $\lambda \times OIB$: the product of absolute OIB and the inverse of a stock's market capitalization in the fourth model and the product of absolute OIB and $PQSPR$ in the fifth model. The coefficients of both of these scaled versions of absolute order imbalance are positive and significant, but the effect of absolute $\lambda \times OIB$ survives. In the final two models of Table 7, we again include the one-day lagged maximum of *Good PIN* and *Bad PIN*. This variable does have a significantly positive coefficient in the final model (in line with expectations), but the coefficient on our absolute private information measure is still positive and significant.

We conclude that, consistent with the prediction of theoretical models of informed trading, our private information measure predicts future return volatility.

4.5. Private information shocks around M&A announcements

To complement our analyses of the relation between our private information measure and firm characteristics, stock returns, and future return volatility discussed thus far, in this section, we further evaluate the validity of our measure as a proxy for private information by studying its behavior around important corporate information events. We focus on target firms in corporate mergers and acquisitions, since M&A transactions tend to have a large impact on the target firm's stock price (e.g., Jensen and Ruback, 1983) and involve a high degree of asymmetric information that generates considerable scope for informed trading (e.g., Keown and Pinkerton, 1981; Meulbroek, 1992).[21] Prior studies also use M&A announcements to assess the validity of various measures of informed trading (e.g., Aktas, de Bodt, Declerck, and Oppens, 2007; Augustin, Brenner, and Subrahmanyam, 2015; Duarte, Hu, and Young, 2017).

We obtain data on all M&A transactions that involve a publicly traded target firm that is part of our sample of NYSE stocks over 2001 to 2014 from Securities Data Company (SDC). Following Hackbarth and Morellec (2008), we only include transactions in which the percentage of ownership of the acquiring firm after the transaction was greater than 50%. Following Augustin et al. (2015), we drop transactions with pending or un-

[21] We do not study earnings announcements since they tend to be much less material in terms of stock price response and since Sarkar and Schwartz (2009) show that trading around M&A announcements is characterized by private information, while trading around earnings announcements is better characterized by belief heterogeneity, which develops much more gradually.

known status and transactions that were later withdrawn, and we restrict the sample to transactions above \$1m. Our final sample consists of 350 M&A transactions.

To perform an event study of the target stock price response around the announcement of the M&A transaction, we map the CUSIP from SDC to the PERMNO from CRSP. Following Johnson and So (2018), we use an event window of [-10,+10] from ten days before till ten days after the announcement. We estimate cumulative abnormal returns (CARs) based on mean-adjusted returns with an estimation window of [-41,-11]. In our main analyses, we limit the sample to the 174 M&A transactions in which the target experienced a significantly positive run-up as indicated by the CAR during the pre-event window [-10,-1], which indicates that significant information was incorporated before the announcement, which is suggestive of informed trading. We obtain broadly similar results for the sample of all transactions and for the sample of transactions with a significantly positive CAR over the full event window [-10,+10]. We also obtain similar results with an event window of [-30,+30] and an estimation window of [-90,-31].

Figure 3 shows four graphs depicting the behavior of, respectively, the cumulative abnormal return, the λ, the OIB, and our simplified private information measure $\lambda \times OIB$ of the target firm for each of the days in the event window [-10,+10]. Consistent with prior studies, the first graph shows that the CAR of the target firm is large and positive, on average close to 25% over the full event window. The pre-event price run-up is around 8%, suggesting that there was considerable scope for private information before these M&A transactions.[22]

The graph of the behavior of the daily λ in the event window [-10,+10] in the second graph of Figure 3 indicates that the estimated price impact parameter for the target

[22]These results are remarkably similar to those of Keown and Pinkerton (1981), who report a cumulative abnormal return of around 28% and a pre-event run-up of around 12% in a sample of 194 M&A transactions in the period 1975-1978, which they interpret as evidence as trading on private information ahead of M&A announcements.

stock is large at the start of the pre-event window (at around 80 basis points for a trade of \$1m.) and then decreases gradually over the pre-event window (possibly as a result of information slowly being incorporated into the stock price ahead of the announcement), before showing a large drop on the event day, followed by a smaller drop on the next day, after which λ is relatively stable at around 15 basis points for a trade of \$1m. This pattern is consistent with a substantial degree of asymmetric information before the event that is largely resolved once the M&A announcement has been made.

This pattern also leads us to reconsider our way to estimate $\lambda \times OIB$ for this application. In our more general analysis of the relation of this private information measure with stock returns or volatility, we use the moving average of a stock's daily λ over the past 20 days as a proxy for that stock's expected price impact on the day of interest that may be taken into account by investors optimally trading on information or liquidity shocks in the cross-section of stocks. This approach seems reasonable in general, since λ tends to be highly persistent. However, in the context of looking at specific information events (or other situations in which λ may change considerably day-by-day), this approach may be less sensible. In particular, investors should anticipate that λ drops after M&A announcements and would thus not use a 20-day moving average to estimate expected price impact in the post-event window. To resolve this issue, we use the estimated λ on the day of interest as an input for our private information measure in our analyses of M&A transactions. We obtain similar results when we take one-day lagged λ.

Consistent with Sarkar and Schwartz (2009), the third graph of Figure 3 shows clear evidence of one-sided trading before the event. OIB is large and positive on every day in the pre-event window (at around \$2m. per day), and even on the day of and the first day after the M&A announcement, possibly because the stock price has only fully adjusted to the information by the end of day one – as indicated by the CAR graph in Figure 3. OIB is large and negative especially on the few days immediately after the announcement, possibly due to the unwinding of informed trades, the exit of corporate insiders, or the desire to not be exposed to deal failure risk (Pedersen, 2015).

The fourth graph of Figure 3 shows that the simplified private information measure $\lambda \times OIB$ assumes large positive values in the pre-event window [-10,-1] as well as on the event day and – to a lesser extent – the day after the event. The former finding is consistent with informed trading based on considerable private information before the event. The latter finding could again be explained by prices only fully adjusting to the new information by the end of day one, which suggests there may still be scope for private information until then. After day one, $\lambda \times OIB$ is close to zero, indicating very small or zero private information shocks.

The results in Figure 3 illustrate the key benefit of incorporating both the amount (OIB) and the costs (λ) of trading in a measure of private information. λ by itself is an ex ante measure of private information, and does not incorporate the actual amount of trading. On the other hand, using OIB as a proxy for informed trading would also lead to misleading results, for example as it would suggest a significant amount of informed trading based on negative information shocks after the event. The combined measure $\lambda \times OIB$ identifies pre-event OIB as an indication of informed trading (since trading is expensive) and post-event OIB primarily as liquidity-motivated trading (since trading is relatively cheap).

In Table 8, we further examine whether the patterns we observe in $\lambda \times OIB$ around M&A announcements are consistent with this measure capturing private information not only in the time-series within the event window as in Figure 3, but also in the cross-section of M&A transactions. In particular, Table 8 shows the results of ten different cross-sectional regressions of the CAR starting at day -10 up to and including day -1 on the cumulative $\lambda \times OIB$ over the same window (based on the full sample of M&A targets). In other words, the first column of Table 8 presents the results of a cross-sectional regression of the abnormal return on day -10 on $\lambda \times OIB$ on day -10, while the final column presents the results of a regression of the cumulative abnormal return over the pre-event window [-10,-1] on the cumulative $\lambda \times OIB$ over [-10,-1]. Our interpretation of $\lambda \times OIB$ as a private information measure predicts a positive coefficient in these regressions, which would indicate that M&A transactions with a greater price run-up tend to have higher values for the private information measure.

Table 8 shows that the coefficient on $\lambda \times OIB$ is positive in all ten regressions, and significantly so in eight out of ten regressions. These effects obtain even after controlling for average λ and cumulative OIB over the same window. The R^2s indicate that a substantial 5% to 20% of the cross-sectional variation in the target run-up can be explained in these regressions.

Taken together, the results in Figure 3 and Table 8 provide further support for the use of $\lambda \times OIB$ as a proxy for the private information shock underlying informed trading in individual stocks. The measure increases for target firms in M&A transactions in the period before the announcement and is greater for target firms whose stock price run-up is suggestive of a greater degree of pre-event private information.

4.6. Private information shocks around the exogenous loss of analyst coverage

We proceed to study the behavior of our private information measure around a set of events that prior studies interpret as exogenous shocks to the degree of asymmetric information for individual stocks: the termination of analyst coverage of a number of stocks due to the closing by 43 U.S. brokerage firms of their research departments between 2000 and 2008. Kelly and Ljungqvist (2012) argue that the closures of these research departments were due to changes in the economics of producing research, and plausibly exogenous to asymmetric information about the firm. They find empirically that the loss of analyst coverage was indeed associated with an increase in information asymmetry. We conjecture that this exogenous decrease in the production of public information resulted in a greater scope for private information.[23]

To test this conjecture and to examine whether any such effects are picked up by our new private information measures, we carry out a difference-in-differences (DiD) analysis of the impact of the loss of analyst coverage on $|v|$ (the absolute value of our private information measure), $\lambda \times |OIB|$ (the absolute simplified private information measure), as well as the latter's components λ and $|OIB|$. Treated stocks are stocks that lost analyst coverage due to the closures of these research departments. We run the DiD based on pre- and post-events periods of 90, 180, and 360 days.

The results are in Table 9. The table reports the equally-weighted average λ, $|OIB|$, $\lambda \times |OIB|$, and $|v|$ in the three different pre- and post-event windows as well as the DiD effect and its statistical significance. Consistent with the notion that asymmetric information increased after the event, we find that the λ of the treated stocks increased significantly relative to the control stocks in the 180-day and 360-day event windows. We also find that OIB decreased significantly in absolute terms for treated stocks in all three event windows, which could be indicative of less intense trading by noise traders in an environment of heightened information asymmetry and greater illiquidity (but could also be consistent with other interpretations).

Importantly, for both private information measures, we find that the exogenous loss of analyst coverage resulted in a greater absolute level of private information for treated stocks after the event. These effects are only statistically significant for the 180-day event window and, at 3% of the pre-event average $\lambda \times |OIB|$ and 7% of the pre-event average $|v|$, relatively modest in an economic sense. That said, the lack of significance for the 90-day event window is consistent with the insignificant change in λ over that window,

[23] We are very grateful to Feng (Jack) Jiang for sharing the list of treated and control firms used in Harford, Jiang, Wang, and Xie (2019), who use data similar to Kelly and Ljungqvist (2012).

which could be consistent with the loss of analyst coverage not immediately affecting asymmetric information in the first quarter following the event. And over the longer 360-day window, it is conceivable that the closures of these research departments by a number of brokers was in part compensated by other brokers taking over the coverage and/or other ways the market dealt with alleviating asymmetric information (even though λ remained high).

All in all, we interpret the evidence in Table 9 as consistent with our private information measures picking up an increase in the scope for private information after an exogenous reduction in the production of public information.

5. Conclusion

This paper proposes a new private information measure for individual securities based on a theoretical trade optimization model. Theoretically, the key innovation of our model is that informed traders in our model strategically optimize their trading in the cross-section of securities based on the securities' price impact parameters. When we introduce a benchmark security that is not subject to informed trading, we obtain a simple closed-form expression for the private information shock underlying informed trading.

Empirically, our private information measure is straightforward to estimate for any security over any time period as twice the product of the security's price impact parameter λ and its order imbalance OIB minus the product of the benchmark security's λ and OIB. Furthermore, in contrast to other measures that proxy for private information, our measure also conveys the direction of the private information signal. In cross-sectional applications, our private information measure simplifies even further to a security's $\lambda \times$ *OIB*. This expression reflects the intuition that observed order flow is more likely to be informed when trading is expensive.

We validate our private information measure by estimating it for all NYSE stocks each day based on intraday data over the period 2001-2014. In particular, we find that the measure is greater for smaller firms with higher analyst dispersion and that it is positively related to contemporaneous stock returns. Consistent with the idea that informed trading is associated with permanent price impact, we show that return reversals are significantly stronger following stock-days with little or no private information than following stock-days with large private information shocks. Our measure also predicts return volatility and increases before M&A announcements as well as after the exogenous loss of analyst coverage.

In sum, our paper offers a new private information measure that is both theoretically motivated and easy to estimate for a wide variety of academic and practical applications.

Appendix A. Proofs

Proof of Lemma 1. We solve the investors' trading problem by a standard Lagrangian multiplier technique. We define:

$$L(\underline{o}_i, \mu) = E(\underline{\delta}|\underline{v}_i)'\underline{o}_i - \underline{o}_i'\mathbf{\Lambda}(\underline{o}_i + E(\underline{o}_{-i})) - \mu(Z_i - \underline{\iota}'\underline{o}_i), \tag{A.1}$$

$$= -\underline{o}_i'\mathbf{\Lambda}(\underline{o}_i) - \mu(Z_i - \underline{\iota}'\underline{o}_i), \tag{A.2}$$

since $\underline{v}_i = \underline{0}$ by assumption and $E(\underline{o}_{-i}) = 0$ because $E(Z_i) = 0$ and $\underline{v}_i = \underline{0}$ $\forall i$. The necessary FOCs for optimality are given by:

$$\frac{\partial L(\underline{o}_i, \mu)}{\partial \underline{o}_i} = 0, \qquad \frac{\partial L(\underline{o}_i, \mu)}{\partial \mu} = 0. \tag{A.3}$$

As $L(\underline{o}_i, \mu)$ contains only polynomial terms of at most second order, we can write the FOCs as a system of linear equations and solve it as is shown below. In matrix form, the

FOCs are given by:

$$\begin{bmatrix} -\underline{a}_i \\ Z_i \end{bmatrix} = \begin{bmatrix} -\mathbf{Q}_i & \underline{\iota} \\ \underline{\iota}' & 0 \end{bmatrix} \begin{bmatrix} \underline{o}_i \\ \mu \end{bmatrix}, \tag{A.4}$$

where

$$\mathbf{Q}_i = 2\mathbf{\Lambda}, \qquad \underline{a}_i = \underline{0}. \tag{A.5}$$

Using the partitioned inverse (see Greene, 2000, p. 34), we obtain our solution:

$$\mu = -(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1}\underline{\iota}'\mathbf{Q}_i^{-1})\underline{a}_i + Z_i(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1} = Z_i(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1} \tag{A.6}$$

$$\underline{o}_i = \mathbf{Q}_i^{-1}(\mathbf{I} - \underline{\iota}(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1}\underline{\iota}'\mathbf{Q}_i^{-1})\underline{a}_i + Z_i\mathbf{Q}_i^{-1}\underline{\iota}(\underline{\iota}\mathbf{Q}_i^{-1}\underline{\iota}')^{-1}. \tag{A.7}$$

$$= \mathbf{Q}_i^{-1}\underline{a}_i + \mathbf{Q}_i^{-1}\underline{\iota}\mu = \mathbf{Q}_i^{-1}\underline{\iota}\mu = Z_i(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}\mathbf{\Lambda}^{-1}\underline{\iota}. \tag{A.8}$$

□

Proof of Lemma 2. With information shocks, the FOCs in Eq. (A.4) change to:

$$\begin{bmatrix} -\underline{a}_i^{inf} \\ Z_i \end{bmatrix} = \begin{bmatrix} -\mathbf{Q}_i & \underline{\iota} \\ \underline{\iota}' & 0 \end{bmatrix} \begin{bmatrix} \underline{x}_i \\ \mu \end{bmatrix}, \tag{A.9}$$

where

$$\mathbf{Q}_i = 2\mathbf{\Lambda}, \qquad \underline{a}_i^{inf} = \underline{v}_i, \tag{A.10}$$

since $E(\underline{\delta}_i|\underline{v}_i) = \underline{v}_i$. Using the partitioned inverse as before, we get

$$\mu = -(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1}\underline{\iota}'\mathbf{Q}_i^{-1})\underline{a}_i^{inf}+Z_i(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1} \tag{A.11}$$

$$\underline{o}_i = \mathbf{Q}_i^{-1}(\mathbf{I}-\underline{\iota}(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1}\underline{\iota}'\mathbf{Q}_i^{-1})\underline{a}_i^{inf}+Z_i\mathbf{Q}_i^{-1}\underline{\iota}(\underline{\iota}\mathbf{Q}_i^{-1}\underline{\iota}')^{-1}. \tag{A.12}$$

$$= Z_i(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}\mathbf{\Lambda}^{-1}\underline{\iota}+\frac{1}{2}\mathbf{\Lambda}^{-1}(\mathbf{I}-(\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}_i^{-1})\underline{v}_i. \tag{A.13}$$

□

Proof of Lemma 3. Pre-multiplying Eq. (6) with $\mathbf{\Lambda}$ and rearranging yields Eq. (7). Solving Eq. (7) towards $\underline{v}_i$ yields

$$\underline{v}_i = 2\mathbf{\Lambda}-2(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}(Z_i-\frac{1}{2}\sum_j \lambda_j^{-1}v_{i,j})\underline{\iota}. \tag{A.14}$$

By assumption, $v_{i,1} = 0$. It follows that

$$\lambda_1 o_{i,1} = (\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}(Z_i-\frac{1}{2}\sum_j \lambda_j^{-1}v_{i,j}). \tag{A.15}$$

Substituting Eq. (A.15) into Eq. (A.14) yields Eq. (8). □

Proof of Proposition 1. Summing Eq. (8) over all investors i trivially yields Eq. (9). □

References

Admati, A. R. and P. Pfleiderer (1988). A Theory of Intraday Patterns: Volume and Price Variability. *Review of Financial Studies 1*, 3–40.

Aktas, N., E. de Bodt, F. Declerck, and H. V. Oppens (2007). The PIN Anomaly Around M&A Announcements. *Journal of Financial Markets 10*, 169 – 191.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Augustin, P., M. Brenner, and M. Subrahmanyam (2015). Informed Options Trading Prior to M&A Announcements: Insider Trading. *working paper, McGill University*.

Back, K., K. Crotty, and T. Li (2017). Identifying Information Asymmetry in Securities Markets. *Review of Financial Studies forthcoming*.

Baker, M. and J. C. Stein (2004). Market Liquidity as a Sentiment Indicator. *Journal of Financial Markets 7*, 271–299.

Boehmer, E. and E. K. Kelley (2009). Institutional Investors and the Informational Efficiency of Prices. *Review of Financial Studies 22*, 3563–3594.

Brennan, M. J., S.-W. Huh, and A. Subrahmanyam (2018). High-Frequency Measures of Informed Trading and Corporate Announcements. *The Review of Financial Studies 31*, 2326–2376.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance 52*, 57–82.

Chordia, T., R. Roll, and A. Subrahmanyam (2005). Evidence on the Speed of Convergence to Market Efficiency. *Journal of Financial Economics 76*, 271–292.

Collin-Dufresne, P. and V. Fos (2015). Do prices reveal the presence of informed trading? *The Journal of Finance 70*(4), 1555–1582.

Cox, D. R. and D. R. Peterson (1994). Stock Returns Following Large One-Day Declines: Evidence on Short-Term Reversals and Longer-Term Performance. *Journal of Finance 49*, 255–267.

Dow, J. and R. Rahi (2003). Informed Trading, Investment, and Welfare. *Journal of Business 76*, 439–454.

Duarte, J., E. Hu, and L. Young (2017). Does the PIN Model Mis-Identify Private Information and If So, What Are Our Alternatives? *working paper, Rice University*.

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

Easley, D., R. F. Engle, M. O'Hara, and L. Wu (2008). Time-Varying Arrival Rates of Informed and Uninformed Trades. *Journal of Financial Econometrics 6*, 171–207.

Easley, D., S. Hvidkjaer, and M. O'Hara (2002). Is Information Risk a Determinant of Asset Returns? *Journal of Finance 57*, 2185–2221.

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman (1996). Liquidity, Information, and Infrequently Traded Stocks. *Journal of Finance 51*, 1405–1436.

Easley, D., M. M. López de Prado, and M. O'Hara (2011). The Microstructure of the 'Flash Crash'. *Journal of Portfolio Management 37*, 118–128.

Easley, D., M. M. López de Prado, and M. O'Hara (2012). Flow Toxicity and Liquidity in a High-frequency World. *Review of Financial Studies 25*, 1457–1493.

Easley, D. and M. O'Hara (1987). Price, Trade Size, and Information in Securities Markets. *Journal of Financial Economics 19*, 69 – 90.

Easley, D. and M. O'Hara (2004). Information and the Cost of Capital. *Journal of Finance 59*, 1553–1583.

Edmans, A. (2009). Blockholder Trading, Market Efficiency, and Managerial Myopia. *Journal of Finance 64*, 2481–2513.

Fama, E. F. and K. R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics 33*, 3–56.

Fama, E. F. and J. D. MacBeth (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy 81*, 607–636.

Fishman, M. J. and K. M. Hagerty (1989). Disclosure Decisions by Firms and the Competition for Price Efficiency. *Journal of Finance 44*, 633–646.

Fong, K., W. C. Holden, and C. Trzcinka (2017). What Are the Best Liquidity Proxies for Global Research? *Review of Finance 21*, 1355–1401.

Gao, X. and J. R. Ritter (2010). The Marketing of Seasoned Equity Offerings. *Journal of Financial Economics 97*, 33 – 52.

Glosten, L. and P. Milgrom (1985). Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics 14*, 71–100.

Glosten, L. R. and L. E. Harris (1988). Estimating the Components of the Bid/Ask Spread. *Journal of Financial Economics 21*, 123–142.

Goldstein, I. and A. Guembel (2008). Manipulation and the Allocational Role of Prices. *Review of Economic Studies 75*, 133–164.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009). Do Liquidity Measures Measure Liquidity? *Journal of Financial Economics 92*, 153 – 181.

Greene, W. (2000). *Econometric Analysis*. Prentice Hall, Englewood Cliff.

Grossman, S. J. and J. E. Stiglitz (1980). On the Impossibility of Informationally Efficient Markets. *American Economic Review 70*, 393–408.

Hackbarth, D. and E. Morellec (2008). Stock Returns in Mergers and Acquisitions. *Journal of Finance 63*, 1213–1252.

Harford, J., F. Jiang, R. Wang, and F. Xie (2019). Analyst Career Concerns, Effort Allocation, and Firms' Information Environment. *Review of Financial Studies 32*, 2179–2224.

Hasbrouck, J. (1991). Measuring the Information Content of Stock Trades. *Journal of Finance 46*, 179–207.

Hasbrouck, J. (2009). Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data. *Journal of Finance 64*, 1445–1477.

Holden, C. W. and S. Jacobsen (2014). Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions. *Journal of Finance 69*, 1747–1785.

Jegadeesh, N. (1990). Evidence of Predictable Behavior of Security Returns. *Journal of Finance 45*, 881–898.

Jensen, M. C. and R. S. Ruback (1983). The Market for Corporate Control: The Scientific Evidence. *Journal of Financial Economics 11*, 5 – 50.

Johnson, T. L. and E. C. So (2018). A Simple Multimarket Measure of Information Asymmetry. *Management Science 64*, 1055–1080.

Kahraman, B. and H. E. Tookes (2017). Trader Leverage and Liquidity. *Journal of Finance 72*, 1567–1610.

Kaniel, R. and H. Liu (2006). So what orders do informed traders use? *The Journal of Business 79*(4), 1867–1913.

Kaul, Gautam, Q. L. and N. Stoffman (2008). AIMing at PIN: Order flow, information, and liquidity. *Working paper, Indiana University*.

Kelly, B. and A. Ljungqvist (2012). Testing Asymmetric-Information Asset Pricing Models. *Review of Financial Studies 25*, 1366–1413.

Keown, A. J. and J. M. Pinkerton (1981). Merger Announcements and Insider Trading Activity: An Empirical Investigation. *Journal of Finance 36*, 855–869.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica 53*, 1315–1335.

Kyle, A. S. and A. A. Obizhaeva (2018). Adverse Selection and Liquidity: From Theory to Practice. *Working Paper, University of Maryland*.

Lai, S., L. Ng, and B. Zhang (2014). Does PIN Affect Equity Prices Around the World? *Journal of Financial Economics 114*, 178–195.

Lau, S., L. Ng, and B. Zhang (2012). Information Environment and Equity Risk Premium Volatility Around the World. *Management Science 58*, 1322–1340.

Lee, C. M. and M. J. Ready (1991). Inferring Trade Direction from Intraday Data. *Journal of Finance 46*, 733–746.

Lee, J. and A. S. Kyle (2018). When Are Financial Markets Perfectly Competitive? Working Paper.

Madhavan, A., M. Richardson, and M. Roomans (1997). Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks. *Review of Financial Studies 10*, 1035–1064.

Madhavan, A. and S. Smidt (1991a). A Bayesian Model of Intraday Specialist Pricing. *Journal of Financial Economics 30*, 99 – 134.

Madhavan, A. and S. Smidt (1991b). A Bayesian Model of Intraday Specialist Pricing. *Journal of Financial Economics 30*, 99 – 134.

Manove, M. (1989). The Harm from Insider Trading and Informed Speculation. *Quarterly Journal of Economics 104*, 823–845.

Marshall, B. R., N. H. Nguyen, and N. Visaltanachoti (2011). Commodity Liquidity Measurement and Transaction Costs. *Review of Financial Studies 25*, 599–638.

Meulbroek, L. K. (1992). An Empirical Analysis of Illegal Insider Trading. *Journal of Finance 47*, 1661–1699.

Nagel, S. (2012). Evaporating Liquidity. *Review of Financial Studies 25*, 2005–2039.

Odders-White, E. R. and M. J. Ready (2008). The Probability and Magnitude of Information Events. *Journal of Financial Economics 87*, 227 – 248.

Pedersen, L. (2015). *Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined*. Princeton University Press.

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *Journal of Finance 39*, 1127–1139.

Rösch, D. M., A. Subrahmanyam, and M. A. van Dijk (2017). The Dynamics of Market Efficiency. *Review of Financial Studies 30*, 1151–1187.

Sadka, R. (2006). Momentum and Post-Earnings-Announcement Drift Anomalies: The Role of Liquidity Risk. *Journal of Financial Economics 80*, 309–349.

Sarkar, A. and R. A. Schwartz (2009). Market Sidedness: Insights into Motives for Trade Initiation. *Journal of Finance 64*, 375–423.

Spiegel, M. and A. Subrahmanyam (1992). Informed Speculation and Hedging in a Noncompetitive Securities Market. *Review of Financial Studies 5*, 307–329.

Stoll, H. R. (1978). The Supply of Dealer Services in Securities Markets. *Journal of Finance 33*, 1133–1151.

Table 1 – Cross-sectional summary statistics of time-series averages

This table reports the cross-sectional (across the 1,388 NYSE stocks in the sample) mean, standard deviation, first quartile, median, and third quartile of the time-series average by stock of the daily percentage return from corporate action adjusted end-of-day mid-quotes (*Return*; in %; cross-sectionally winsorized each day at the 0.1% and 99.9% level), the daily percentage average proportional quoted spread ($PQSPR$; in %), the price impact defined as the percentage price change associated with a million US dollars of trading volume (λ; expressed as %Return/US$m.; cross-sectionally winsorized each day at the 95% level), the daily order imbalance or difference between the total US dollar volume of trades initiated by buyers and sellers expressed in millions of dollars (OIB; in US$m.), the daily private information measure v (defined as twice the $\lambda \times OIB$ of the stock of interest minus the $\lambda \times OIB$ of the SPDR (or SPY) benchmark security, from Eq. (9)), the simplified daily private information measure $\lambda \times OIB$, and the Easley et al. (2002) probability of informed trading (PIN) estimated daily into *Good PIN* and *Bad PIN* following Brennan et al. (2018). The first column indicates the number of stocks over which the summary statistics are computed. The second column indicates the average number of days a stock is in the sample. The sample includes 1,388 NYSE stocks during our sample period 2001-2014. Data to compute all variables in the table are from TRTH. The factor to adjust daily closing mid-quote data for corporate actions is from CRSP.

	#stocks	days	mean	stddev	25%	median	75%
Return	1,388	1,992	0.06	0.14	0.03	0.05	0.08
$PQSPR$	1,388	1,992	0.13	0.09	0.08	0.11	0.17
λ	1,388	1,992	0.83	0.87	0.20	0.51	1.13
OIB	1,388	1,992	1.19	2.41	0.03	0.35	1.36
$v = 2(\lambda_i OIB_i - \lambda_{SPY} OIB_{SPY})$	1,388	1,992	0.30	0.62	-0.03	0.19	0.49
$\lambda \times OIB$	1,388	1,992	0.19	0.30	0.04	0.15	0.29
Good PIN	1,260	2,173	0.19	0.04	0.17	0.19	0.22
Bad PIN	1,260	2,173	0.14	0.03	0.12	0.14	0.16

Table 2 – Pooled correlations of daily private information, liquidity, order imbalance, and returns

This table reports pooled Pearson correlation coefficients between nine daily stock-specific variables: the private information measure (v), the absolute private information measure ($|v|$), the simplified private information measure ($\lambda\times OIB$), the absolute simplified private information measure ($\lambda\times|OIB|$), the price impact (λ), the dollar order imbalance (OIB), the proportional quoted spread ($PQSPR$), the dollar trading volume ($Volume$), the stock returns (*Return*), and *Good PIN* and *Bad PIN* as in Brennan et al. (2018). We refer to Table 1 for a description of these variables. The sample includes 1,388 NYSE stocks during our sample period 2001-2014. Data to compute the variables are from TRTH and CRSP. The table reports p-values in parentheses below the correlations.

	v	$\|v\|$	$\lambda\times OIB$	$\lambda\times\|OIB\|$	λ	OIB	$PQSPR$	$Volume$	$Return$	$G.\ PIN$	$B.\ PIN$
$\|v\|$	0.012	1.000									
	(0.00)										
$\lambda\times OIB$	0.935	0.025	1.000								
	(0.00)	(0.00)									
$\lambda\times\|OIB\|$	0.052	0.932	0.064	1.000							
	(0.00)	(0.00)	(0.00)								
λ	-0.042	0.143	-0.031	0.131	1.000						
	(0.00)	(0.00)	(0.00)	(0.00)							
OIB	0.452	0.095	0.489	0.120	-0.063	1.000					
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)						
$PQSPR$	0.011	0.028	0.010	0.026	0.286	-0.017	1.000				
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)					
$Volume$	-0.010	0.112	-0.010	0.121	-0.233	0.025	-0.130	1.000			
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)				
$Return$	0.103	0.080	0.183	0.093	0.012	0.116	-0.003	0.002	1.000		
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)			
$Good\ PIN$	0.230	0.149	0.249	0.175	-0.009	0.183	0.012	0.064	0.102	1.000	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		
$Bad\ PIN$	-0.152	0.066	-0.173	0.065	-0.004	-0.113	0.000	0.074	-0.097	-0.203	1.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.79)	(0.00)	(0.00)	(0.00)	

Table 3 – Characteristics of stocks with large vs. small private information shocks

This table reports the following average characteristics of quintile portfolios sorted on the absolute private information measure ($|v|$) each month: the market capitalization in billions of dollars ($Mktcap$), the squared stock return ($Volatility$), the proportional quoted spread ($PQSPR$), the price impact defined as the percentage price change associated with a million dollars of trading volume (λ), the number of analysts covering the firm (*Analyst coverage*), and the dispersion in analysts' earnings forecasts, defined as the standard deviation of annual analysts' earnings per share forecasts scaled by the mean forecast (*Analyst dispersion*). Portfolios are formed every month and all variables are estimated using data from the previous month. The sample includes 1,388 NYSE stocks during our sample period 2001-2014. Data to compute the variables are from TRTH, CRSP, and I/B/E/S. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

$\|v\|$ Quintiles	$Mktcap$	$Volatility$	$PQSPR$	λ	*Analyst coverage*	*Analyst dispersion*
1	12.03	4.09	0.11	0.61	10.97	0.05
2	10.91	4.48	0.10	0.60	11.17	0.06
3	10.50	5.47	0.11	0.65	11.15	0.07
4	10.55	6.56	0.11	0.71	11.33	0.10
5	10.71	9.15	0.12	0.82	11.83	0.15
Difference [5] - [1]	-1.33***	5.06***	0.01***	0.21***	0.86***	0.10***

Table 4 – Daily Fama-MacBeth regressions of returns on contemporaneous private information

This table reports the time-series averages of the estimated slope coefficients from daily Fama-MacBeth regressions to explain cross-sectional variation in daily stock returns. The dependent variable is the end-of-day mid-quote return of stock i on day d ($Return_{i,d}$). The independent variables are: the return of stock i on day $d-1$ ($Return_{i,d-1}$), the order imbalance of stock i on day d ($OIB_{i,d}$), the price impact of stock i on day d – calculated as the stock's average price impact estimate over the past 20 days with setting non-positive price impact estimates to zero ($\lambda_{i,d}$), the inverse of the market capitalization of stock i at the beginning of each year ($1/Mktcap_{i,y-}$), the proportional quoted spread of stock i on day $d-1$ ($PQSPR_{i,d-1}$), *Good PIN* and *Bad PIN* as in Brennan et al. (2018), as well as various interaction terms. The table reports Fama-MacBeth t-statistics with Newey-West corrections in parentheses below the average coefficients. Some coefficients have been scaled for ease of presentation. Intercepts are suppressed to conserve space. The final two rows report the R^2 and the number of regressions. The sample includes 1,388 NYSE stocks during our sample period 2001-2014. Data to compute the variables are from TRTH and CRSP. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable: $Return_{i,d}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$Return_{i,d-1}$	-0.020***	-0.020***	-0.021***	-0.021***	-0.022***	-0.024***
	(-9.21)	(-9.27)	(-9.83)	(-9.66)	(-10.34)	(-11.16)
$\lambda_{i,d}\times OIB_{i,d}$	0.323***	0.306***	0.282***	0.283***	0.287***	0.250***
	(49.55)	(45.66)	(41.51)	(42.16)	(45.55)	(38.25)
$\lambda_{i,d}\ \times 10^2$		0.057***	0.046***	0.050***	0.054***	0.043***
		(9.26)	(6.81)	(8.54)	(8.67)	(6.92)
$OIB_{i,d}\ \times 10^4$		0.317***	0.323***	-0.382***	0.282***	-0.336***
		(8.02)	(8.39)	(-6.86)	(7.27)	(-6.30)
$1/Mktcap_{i,y-}$			0.110			0.068
			(2.28)			(1.47)
$OIB_{i,d}\times 1/Mktcap_{i,y-}$			0.275***			0.200***
			(7.18)			(5.43)
$PQSPR_{i,d-1}$				0.001**		0.001*
				(2.40)		(1.67)
$OIB_{i,d}\times PQSPR_{i,d-1}$				0.002***		0.002***
				(14.51)		(13.97)
$Good\ PIN_{i,d}$					0.002***	0.002***
					(18.71)	(17.80)
$Bad\ PIN_{i,d}$					-0.001***	-0.001***
					(-9.10)	(-8.50)
R^2	0.08	0.10	0.11	0.10	0.11	0.13
# regressions	3,415	3,415	3,415	3,415	3,412	3,412

Table 5 – Daily Fama-MacBeth regressions of returns on previous day private information

This table reports the time-series averages of the estimated slope coefficients from daily Fama-MacBeth regressions to explain cross-sectional variation in daily stock returns. The dependent variable is the end-of-day mid-quote return of stock i on day d ($Return_{i,d}$). The independent variables are: the return of stock i on day $d-1$ ($Return_{i,d-1}$), the previous day absolute simplified private information measure $\lambda\times OIB$ computed as the product of the price impact of stock i on day $d-1$ – which is calculated as the stock's average price impact estimate over the past 20 days with setting non-positive price impact estimates to zero ($\lambda_{i,d-1}$) and the absolute order imbalance of stock i on day $d-1$ ($|OIB_{i,d-1}|$), $\lambda_{i,d-1}$ and $|OIB_{i,d-1}|$ separately, as well as various interaction terms. In the last two specifications we further include the maximum of *Good PIN* and *Bad PIN* (estimated as in Brennan et al., 2018) of stock i on day $d-1$. The table reports Fama-MacBeth *t*-statistics with Newey-West corrections in parentheses below the average coefficients. Some coefficients have been scaled for ease of presentation. Intercepts are suppressed to conserve space. The final three rows report the R^2, the adjusted R^2, and the number of regressions. The sample includes 1,388 NYSE stocks during our sample period 2001-2014. Data to compute the variables are from TRTH and CRSP. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable: $Return_{i,d}$					
	(1)	(2)	(3)	(4)	(5)	6
$Return_{i,d-1}$	-0.008***	-0.014***	-0.002	-0.004	-0.013**	-0.000
	(-3.66)	(-5.23)	(-0.65)	(-1.33)	(-3.62)	(-0.05)
$\lambda_{i,d-1}\times\|OIB_{i,d-1}\|$		1.630***		2.131***	1.679***	2.169***
		(4.87)		(5.81)	(4.94)	(5.80)
$Return_{i,d-1}\times\lambda_{i,d-1}\times\|OIB_{i,d-1}\|$		33.547***		35.403***	36.030***	42.16***
		(3.75)		(3.43)	(3.90)	(3.92)
$\lambda_{i,d-1}\ \times 10^2$			0.026***	0.022***		0.022***
			(3.79)	(2.93)		(2.90)
$Return_{i,d-1}\times\lambda_{i,d-1}$			-0.015***	-0.019***		-0.019***
			(-5.88)	(-6.82)		(-7.07)
$\|OIB_{i,d-1}\|\ \times 10^4$			-0.028	-0.117***		-0.116***
			(-1.10)	(-3.90)		(-3.93)
$Return_{i,d-1}\times\|OIB_{i,d-1}\|$			3.837***	0.232		-0.046
			(2.88)	(0.16)		(-0.03)
$max(Good\ PIN_{i,d-1}, Bad\ PIN_{i,d-1})$					0.0001	0.0001
					(1.55)	(1.59)
$Return_{i,d-1}\times max(Good\ PIN_{i,d-1}, Bad\ PIN_{i,d-1})$					-0.003	-0.007**
					(-1.05)	(-2.35)
R^2	0.01	0.02	0.04	0.05	0.03	0.05
Adj. R^2	0.0133	0.0207	0.0321	0.0385	0.0253	0.0428
# regressions	3,415	3,415	3,415	3,415	3,411	3,411

Table 6 – Alphas of reversal strategies among stocks with large vs. small private information shocks

This table reports the results of time-series regressions of factor models to explain returns on a daily reversal strategy within different quintile portfolios sorted on absolute private information. We first sort all stocks in our sample into quintile portfolios based on their $\lambda\times|OIB|$ on day $d-1$. Subsequently, we sort stocks within each $\lambda\times|OIB|$ quintile into five subportfolios based on their returns on day $d-1$. We then compute the returns on a simple reversal strategy within each $\lambda\times|OIB|$ quintile that is long in day $d-1$'s loser stocks (subportfolio 1) and short in day $d-1$'s winner stocks (subportfolio 5) in that quintile. The returns of the reversal strategy are based on these stocks' next day's returns computed from the market close on day $d-1$ till the market close on day d. In the first four columns of the table, the dependent variable is the daily return on the reversal strategy for private information Quintile 1 (small absolute private information shocks). In the fifth column, the dependent variable is the return on the reversal strategy for private information Quintile 5 (large absolute private information shocks). In the final column, the dependent variable is difference in the reversal strategy returns between Quintiles 1 and 5. As dependent variables, we include a number of widely used asset pricing factors: the daily market excess return ($Mkt-RF$), the daily return difference between small and large stocks (SMB), the daily return difference between high and low book-to-market stocks (HML), the daily return difference between past medium-term winner and loser stocks ($Momentum$), the daily return difference between past short-term loser and winner stocks ($Reversal$). The table reports Newey-West t-statistics in parentheses below the coefficients. The final two rows report the R^2 and the number of observations. The sample includes 1,388 NYSE stocks during our sample period 2001-2014. Data to compute the variables are from TRTH and CRSP. Daily factor portfolio returns are from the website of Ken French. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Reversal strategy returns for:	Quintile 1	Quintile 1	Quintile 1	Quintile 1	Quintile 5	Quintile 1 - Quintile 5
Alpha	0.076***	0.077***	0.075***	0.070***	0.007	0.063***
	(4.88)	(4.91)	(4.78)	(4.46)	(0.34)	(3.59)
Mkt - RF	0.111***	0.116***	0.142***	0.109***	0.113***	-0.004
	(3.84)	(3.89)	(4.65)	(3.93)	(3.67)	(-0.17)
SMB		-0.019	-0.037	-0.022	-0.036	0.014
		(-0.41)	(-0.74)	(-0.47)	(-0.64)	(0.29)
HML		-0.031	-0.001	0.018	-0.015	0.033
		(-0.65)	(-0.03)	(0.38)	(-0.21)	(0.62)
Momentum			0.084***	0.089***	0.124***	-0.035
			(2.71)	(2.79)	(3.00)	(-1.14)
Reversal				0.126***	0.230***	-0.104***
				(3.32)	(4.54)	(-2.71)
R^2	0.02	0.02	0.03	0.04	0.05	0.01
# Obs.	3,415	3,415	3,415	3,415	3,415	3,415

Table 7 – Daily Fama-MacBeth regressions of volatility on previous day private information

This table reports the time-series averages of the estimated slope coefficients from daily Fama-MacBeth regressions to explain cross-sectional variation in daily stock return volatility. The dependent variable is the squared end-of-day mid-quote return of stock i on day d ($Return^2_{i,d}$). The independent variables are: the return on day $d-1$ ($Return_{i,d-1}$), the squared return on days $d-1$ through day $d-5$ ($Return^2_{i,d-x}$), the absolute order imbalance on day $d-1$ ($|OIB_{i,d-1}|$), the price impact on day $d-1$ – calculated as the stock's average price impact estimate over the past 20 days with setting non-positive price impact estimates to zero ($\lambda_{i,d-1}$), the inverse of the market capitalization at the beginning of each year ($1/Mktcap_{i,y-}$), the proportional quoted spread on day $d-1$ ($PQSPR_{i,d-1}$), and various interaction terms. In the last two specifications we further include the maximum of *Good PIN* and *Bad PIN* (estimated as in Brennan et al., 2018) of stock i on day $d-1$. The table reports Fama-MacBeth t-statistics with Newey-West corrections in parentheses below the average coefficients. Intercepts are suppressed to conserve space. The final two rows report the R^2 and the number of regressions. The sample includes 1,388 NYSE stocks during our sample period 2001-2014. Data to compute the variables are from TRTH and CRSP. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable: $Return^2_{i,d}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Return_{i,d-1}$	-0.000	-0.000	-0.001	-0.001	-0.001	-0.001	-0.001
	(-0.52)	(-0.46)	(-0.92)	(-0.80)	(-0.94)	(-0.88)	(-0.98)
$Return^2_{i,d-1}$	0.125***	0.103***	0.088***	0.086***	0.083***	0.084***	0.073***
	(17.16)	(15.50)	(13.04)	(12.46)	(12.17)	(12.19)	(10.58)
$Return^2_{i,d-2}$		0.071***	0.062***	0.060***	0.060***	0.061***	0.051***
		(11.97)	(10.82)	(9.96)	(10.58)	(10.28)	(8.71)
$Return^2_{i,d-3}$		0.048***	0.041***	0.040***	0.038***	0.040***	0.032***
		(17.13)	(14.17)	(13.19)	(13.35)	(13.93)	(10.99)
$Return^2_{i,d-4}$		0.050***	0.042***	0.042***	0.041***	0.043***	0.039***
		(9.87)	(8.97)	(8.77)	(8.72)	(9.25)	(8.68)
$Return^2_{i,d-5}$		0.049***	0.043***	0.042***	0.042***	0.043***	0.037***
		(14.08)	(12.80)	(12.07)	(12.27)	(12.70)	(11.33)
$\lambda_{i,d-1} \times \lvert OIB_{i,d-1} \rvert$			0.617***	0.260***	0.515***	0.574***	0.186***
			(10.50)	(2.74)	(7.92)	(9.97)	(2.75)
$\lambda_{i,d-1}$			0.009***	0.007***	0.001	0.009***	0.002*
			(9.27)	(4.96)	(1.05)	(9.09)	(1.73)
$\lvert OIB_{i,d-1} \rvert$			-0.006	0.001	-0.041**	-0.007	-0.030
			(-0.56)	(0.05)	(-2.17)	(-0.62)	(-1.60)
$1/Mktcap_{i,y-}$				0.035***			-0.010
				(3.45)			(-0.77)
$\lvert OIB_{i,d-1} \rvert \times 1/Mktcap_{i,y-}$				172.26**			0.063***
				(2.07)			(5.95)
$PQSPR_{i,d-1}$					0.001***		0.001
					(13.43)		(14.05)
$\lvert OIB_{i,d-1} \rvert \times PQSPR_{i,d-1}$					1.131**		1.02**
					(2.44)		(2.25)
$max(Good\ PIN_{i,d-1}, Bad\ PIN_{i,d-1})$						0.00002	0.00003**
						(1.22)	(2.00)
R^2	0.02	0.05	0.07	0.07	0.07	0.07	0.12
# regressions	3,407	3,407	3,407	3,407	3,407	3,407	3,407

Table 8 – Cross-sectional regressions of target run-up in M&A transactions on private information

This table reports the results of ten different cross-sectional regressions of the cumulative abnormal stock returns (CARs) of target firms in the windows [-10,-10] through [-10,-1] before M&A announcements (the target run-up) on our private information measure. The announcement (event) date for target firm i is denoted by e_i. Daily abnormal returns are calculated as the daily stock return (from corporate action adjusted end-of-day mid-quotes) in excess of the mean return over the estimation window from day e_i-41 through day e_i-11. The dependent variable in column d (d = -10, ..., -1) is the CAR of stock i from day e_i-10 through day e_i+d ($CAR_{i,e_i-10:e_i+d}$). The independent variables are: the cumulative simplified private information measure $\lambda\times OIB$ for stock i from day e_i-10 through day e_i+d, the average price impact parameter of stock i from day e_i-10 through day e_i+d, and the cumulative order imbalance of stock i from day e_i-10 through day e_i+d. The table reports Newey-West t-statistics in parentheses below the coefficients. Intercepts are suppressed to conserve space. The final two rows report the R^2 and the number of observations. (We note that the number of observations varies slightly across the regressions because of some missing end-of-day mid-quotes.) The sample includes all 350 M&A transactions from Securities Data Company (SDC) that involve a publicly traded target firm that is part of our sample of NYSE stocks over 2001 to 2014, in which the percentage of ownership of the acquiring firm after the transaction was greater than 50%, and with a deal value above $1m. Data to compute the variables are from TRTH and CRSP. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable: $CAR_{i,e_i-10:e_i+d}$									
	$d=-10$	-9	-8	-7	-6	-5	-4	-3	-2	-1
$\sum_{j=-10}^{d}(\lambda\times OIB)_{i,e_i+j}$	0.662***	0.355*	0.374	0.372	0.329*	0.386**	0.268*	0.285**	0.504***	0.494**
	(2.72)	(1.65)	(1.44)	(1.62)	(1.70)	(2.18)	(1.83)	(2.04)	(2.77)	(2.46)
$\frac{1}{11+d}\sum_{j=-10}^{d}\lambda_{i,e_i+j}$	0.331	0.601***	1.042**	1.120**	0.939*	1.496**	1.721**	1.663**	1.820**	1.267
	(1.44)	(2.66)	(2.23)	(2.37)	(1.93)	(2.43)	(2.44)	(2.27)	(2.06)	(1.23)
$\sum_{j=-10}^{d} OIB_{i,e_i+j}$	0.009	0.062**	0.093**	0.064**	0.052**	0.027	0.042**	0.041**	0.035**	0.018
	(0.50)	(2.22)	(2.58)	(2.30)	(2.02)	(0.84)	(2.39)	(2.50)	(2.35)	(1.00)
R^2	0.15	0.15	0.21	0.17	0.11	0.06	0.08	0.08	0.11	0.06
# *Obs.*	331	330	331	331	334	334	336	337	336	335

Table 9 – Difference-in-differences of private information around the exogenous loss of analyst coverage

This table reports the equally-weighted average price impact (λ), absolute dollar order imbalance expressed in billions of US dollars ($|OIB|$), absolute simplified private information measure ($\lambda \times |OIB|$), and absolute private information measure ($|v|$), across all treated and control stocks in windows of 90, 180, and 360 days before and after the exogenous loss of analyst coverage due to the closing of the research departments of a number of U.S. brokerage firms. We refer to Table 1 for a description of these variables. Treated stocks are stocks that lost analyst coverage due to the event. Treated and control stocks are from Harford et al. (2019) and also need to be part of our sample, which includes 1,388 NYSE stocks during our sample period of 2001-2014. The last two columns report the average difference-in-differences (DiD) coefficient for each of the four variables and for each of the three event windows, and the respective p-value. Data to compute the variables are from TRTH and CRSP. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

		Treated stocks		Control stocks		Mean	p-value
	Window	Before	After	Before	After	DiD	DiD = 0
λ	90 days	0.437	0.542	0.431	0.519	0.018	0.135
	180 days	0.448	0.609	0.430	0.561	0.029***	0.001
	360 days	0.460	0.613	0.444	0.565	0.032***	0.000
$\|OIB\|$	90 days	0.614	0.560	0.586	0.599	-0.068***	0.002
	180 days	0.606	0.549	0.603	0.605	-0.060***	0.000
	360 days	0.619	0.523	0.591	0.567	-0.072***	0.000
$\lambda \times \|OIB\|$	90 days	0.089	0.100	0.079	0.091	0.000	0.851
	180 days	0.093	0.111	0.081	0.096	0.003*	0.082
	360 days	0.096	0.106	0.083	0.094	-0.001	0.433
$\|v\|$	90 days	1.158	1.363	0.976	1.151	0.030	0.565
	180 days	1.243	1.515	1.021	1.195	0.098**	0.016
	360 days	1.319	1.381	1.083	1.113	0.032	0.272

Figure 1 – Dynamics of private information measure over time

This figure shows the monthly time-variation in the equally-weighted, aggregate private information measure (v) of the 10% of all stocks with the highest and lowest private information on each given day, where the monthly v is computed as the equally-weighted average across days within the month. Data to compute v are from TRTH.

Figure 2 – Time-series of the average return of the top and bottom 10% of all stocks sorted by v

This figure shows the monthly time-variation in the equally-weighted, mid-quote returns of the 10% of all stocks with the highest and lowest private information (v) on each given day, where the monthly average return is computed as the equally-weighted average across days within the month. Data to compute the variables are from TRTH and CRSP.

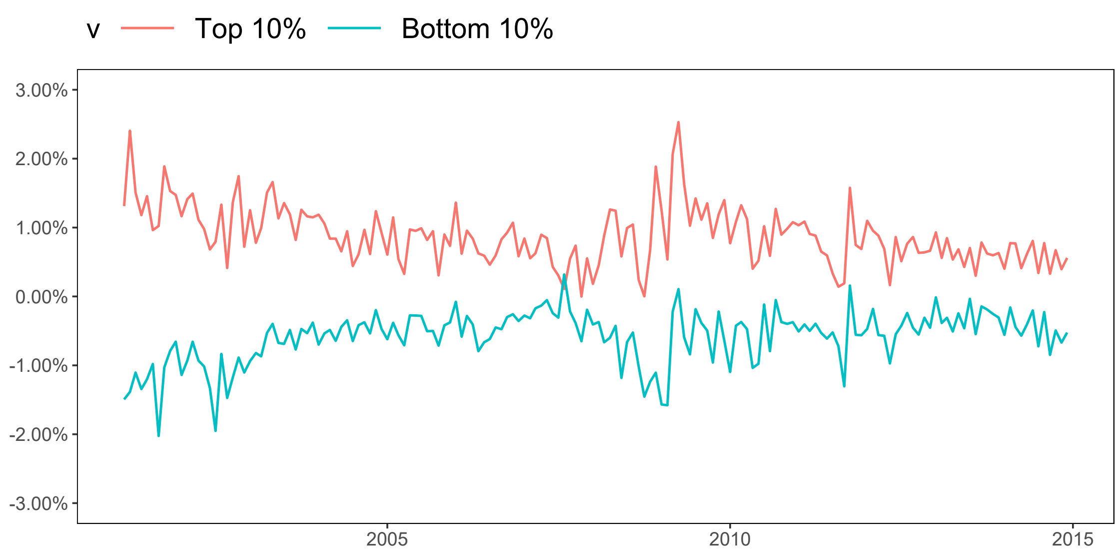
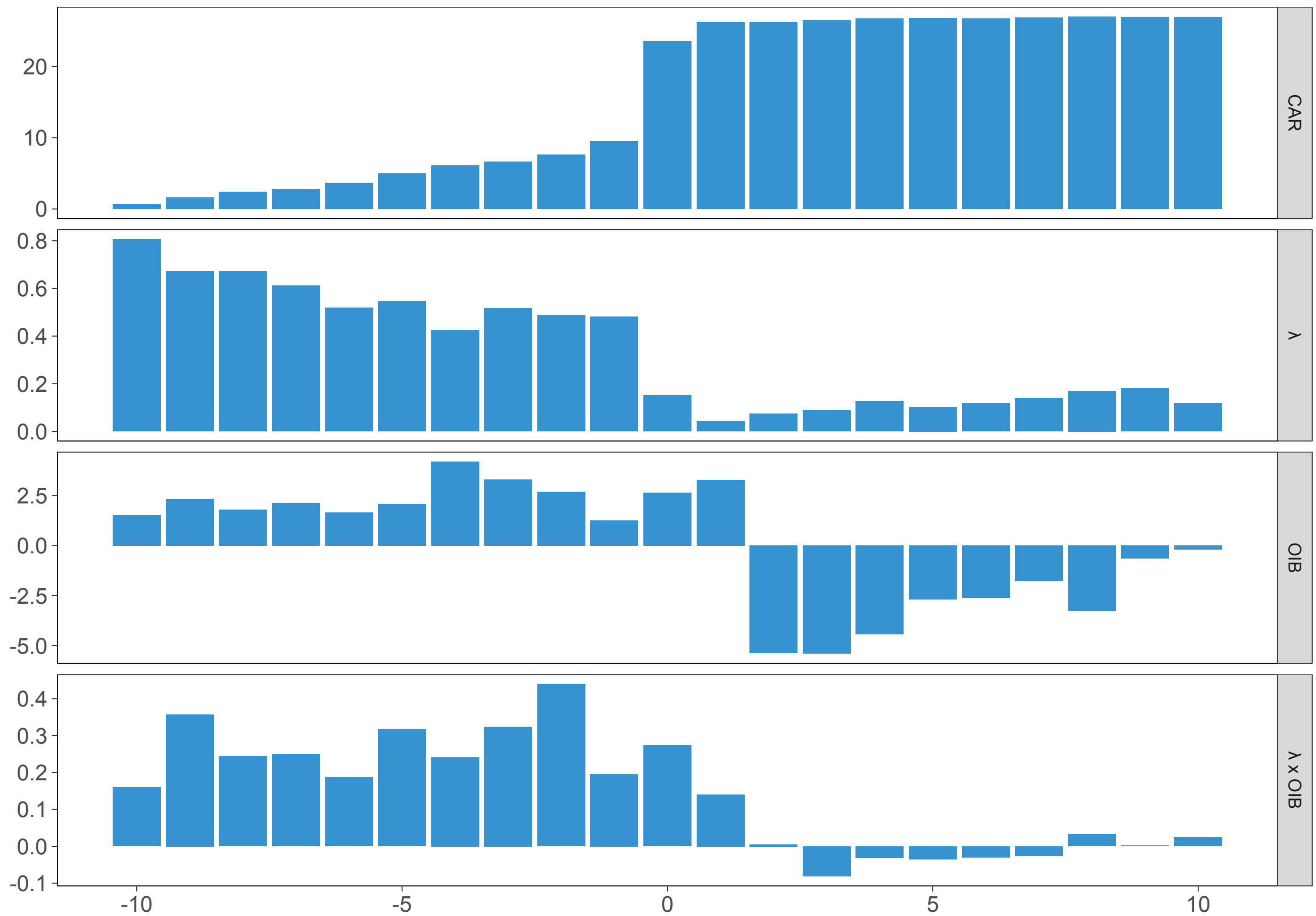


Figure 3 – Event study of private information around M&A announcements

This figure shows the daily cumulative abnormal return (CAR, in %), price impact (λ, in % per million dollars of trading volume), order imbalance (OIB, in millions of dollars), and simplified private information measure ($\lambda \times OIB$) of target firms in the event window [-10,+10] around M&A announcements. The sample includes all 174 M&A transactions from Securities Data Company (SDC) that involve a publicly traded target firm that is part of our sample of NYSE stocks over 2001 to 2014, in which the percentage of ownership of the acquiring firm after the transaction was greater than 50%, with a deal value above $1m, and with a significantly positive run-up as measured by the CAR over the pre-event window [-10,-1]. Data to compute the variables are from TRTH and CRSP.



Internet Appendix

accompanying the paper:

"A strategic trade optimization approach to identifying private information"

by Dion Bongaerts, Dominik Rösch, and Mathijs van Dijk

Internet Appendix IA.1. Overview of notation used

Parameters		
Symbol	*Support*	*Description*
$\underline{\delta}$	$\mathbb{R}^N$	Fundamental mispricing
λ_j	$\mathbb{R}^+$	Price impact parameter of security j
$\mathbf{\Lambda}$	$\mathbb{R}^{N+} \times \mathbb{R}^{N+}$	Diagonal matrix containing all λ_js
Z_i	$\mathbb{R}$	Liquidity shock of investor i
$\underline{v}_i$	$\mathbb{R}^N$	Private information shocks for investor i
Indices		
i	$\{1, .., M\}$	Investors
j	$\{1, .., N\}$	Securities
Decision variables		
$\underline{o}_i$	$\mathbb{R}^N$	Order flow of investor i

Internet Appendix IA.2. Extensions and robustness of the model

In this Appendix, we present several variations of our baseline model (featuring a single informed investor with a perfect signal) that all yield the same basic result: we can use price impact parameters and order flows to reverse-engineer private information shocks, irrespective of whether private information signals are noisy or not or whether only one or multiple investors receive signals. In order to conduct these additional analyses, we need to make distributional assumptions about the noise process $u_{i,j}$, since the signal to noise ratio determines the degree to which the informed investor discounts her signal. These distributional assumptions vary across the variations of the baseline model presented.

Internet Appendix IA.2.1. Noisy signal with only one informed investor

In the first variation of the baseline model, there is still only one informed investor, but the signal that this investor receives is noisy. As a result, the informed investor discounts the signal and trades on the discounted signal. As econometricians we can still reverse-engineer the discounted signal (which is equally if not more informative than the signal itself).

Assume that the noise $u_{i,j}$ is distributed according to $N(0, \sigma^2_{u_j})$ and that the fundamental under-valuation δ_j is distributed according to $N(0, \sigma^2_{F_j})$. We then obtain:[IA.1]

$$E(\delta_j | v_{i,j}) = \frac{Cov(\delta_j, \delta_j + u_{i,j})}{Var(v_{i,j})} v_{i,j} = \phi_j v_{i,j} \equiv \tilde{v}_{i,j}, \tag{IA.1}$$

where

$$\phi_j = \frac{\sigma^2_{F_j}}{\sigma^2_{F_j} + \sigma^2_{u_j}}. \tag{IA.2}$$

[IA.1] $E(\delta_j | v_{i,j})$ is essentially the fitted value of a regression of δ_j on $v_{i,j}$.

As econometricians, we are ultimately interested in $\tilde{\underline{v}}_i$ rather than $\underline{v}_i$. $\underline{v}_i$ is the signal, while $\tilde{\underline{v}}_i$ is the information extracted from the signal by discounting it due to the presence of noise. Similarly, in forming their demand, investors simply include $\tilde{\underline{v}}_i$ rather than $\underline{v}_i$ in Eq. (6) to (9). Hence, in the presence of noisy signals, we can still use our framework to reverse-engineer private information shocks:

Corollary 1. *In the presence of signal noise and a single informed investor per security, the best estimate of order flow-implied private information is given by:*

$$\tilde{\underline{v}} = 2\mathbf{\Lambda}\underline{o} - 2\lambda_1 o_1 \underline{\iota}. \tag{IA.3}$$

Proof. We have that $E(\underline{\delta}|\underline{v}_i) = \tilde{\underline{v}}_i$. Hence, the Lagrangian is equal to Eq. (A.9), but with $a_i^{inf} = \tilde{\underline{v}}_i$. It immediately follows from Lemma 2 and Proposition 1 that Eq. (IA.3) must hold. □

Internet Appendix IA.2.2. Perfect signal with multiple informed investors

Now consider the case with multiple investors that receive the same, perfectly accurate signal about a specific security j. This case is relevant as, in general, informed investors are expected to trade in the same direction. As investors are strategic and rational, they would endogenize the demand of their competitors into their own security demand. The demand by their competitors is damaging for them as it gives rise to additional price impact and therefore elevates transaction costs.

We assume the number of informed investors to be ex ante known to equal k. We also assume k to be the same across securities.[IA.2] Each informed investor i knows that there are $k-1$ other informed investors. As a result, $E(\underline{o}_{-i}) \neq 0$ for each informed investor i

[IA.2] Note that this does not imply that private information is equally important for all securities as $var(\delta_j)$ may differ across securities j.

and the demand is adjusted downwards accordingly:

Lemma 4. *In the presence of k perfectly informed investors, optimal order flow for informed investor i is given by*

$$\underline{o}_i = Z_i(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}\mathbf{\Lambda}^{-1}\underline{\iota}+\frac{1}{k+1}\mathbf{\Lambda}^{-1}(\mathbf{I}-(\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota\iota}'\mathbf{\Lambda}_i^{-1})\underline{v}_i. \tag{IA.4}$$

Proof. With k perfect information shocks, the FOCs in Eq. (A.4) change to:

$$\begin{bmatrix} -\underline{a}_i^{mult} \\ \\ Z_i \end{bmatrix} = \begin{bmatrix} -\mathbf{Q}_i & \underline{\iota} \\ \\ \underline{\iota}' & 0 \end{bmatrix} \begin{bmatrix} \underline{x}_i \\ \\ \mu \end{bmatrix}, \tag{IA.5}$$

where

$$\mathbf{Q}_i = 2\mathbf{\Lambda} \tag{IA.6}$$

$$\underline{a}_i^{mult} = \underline{v}_i - \mathbf{\Lambda}E(\underline{o}_{-i}). \tag{IA.7}$$

Using the partitioned inverse as before, we get

$$\mu = -(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1}\underline{\iota}'\mathbf{Q}_i^{-1})\underline{a}_i^{mult}+Z_i(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1} \tag{IA.8}$$

$$\underline{o}_i = \mathbf{Q}_i^{-1}(\mathbf{I}-\underline{\iota}(\underline{\iota}'\mathbf{Q}_i^{-1}\underline{\iota})^{-1}\underline{\iota}'\mathbf{Q}_i^{-1})\underline{a}_i^{mult}+Z_i\mathbf{Q}_i^{-1}\underline{\iota}(\underline{\iota}\mathbf{Q}_i^{-1}\underline{\iota}')^{-1}. \tag{IA.9}$$

$$= Z_i(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}\mathbf{\Lambda}^{-1}\underline{\iota}+\frac{1}{2}\mathbf{\Lambda}^{-1}(\mathbf{I}-(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}\underline{\iota\iota}'\mathbf{\Lambda}^{-1})(\underline{v}_i-\mathbf{\Lambda}E(\underline{o}_{-i})). \tag{IA.10}$$

These FOCs hold for each investor i. Summing over all investors and taking expectations yields

$$E(\underline{o}) = \frac{1}{2}\mathbf{\Lambda}^{-1}(\mathbf{I}-(\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota\iota}'\mathbf{\Lambda}_i^{-1})(k\underline{v}_i-(k-1)E(\underline{o})), \tag{IA.11}$$

for any informed investor i. We have that

$$\mathbf{\Lambda}^{-1}(\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}_i^{-1}\mathbf{\Lambda}E(\underline{o}_i) = (\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}_i^{-1}E(\underline{o}_i) = \underline{0}. \tag{IA.12}$$

In equilibrium, we have that

$$E(\underline{o}) = \frac{k}{k-1}E(\underline{o}_{-i}), \tag{IA.13}$$

for any informed investor i. Substituting Eq. (IA.13) into Eq. (IA.11) and solving yields

$$E(\underline{o}_{-i}) = \frac{k-1}{k+1}\mathbf{\Lambda}^{-1}(\mathbf{I}-(\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}_i^{-1})\underline{v}_i. \tag{IA.14}$$

Substituting Eq. (IA.14) into Eq. (IA.10) and solving towards $\underline{o}_i$ yields

$$\begin{aligned}\underline{o}_i = {} & Z_i(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}\mathbf{\Lambda}^{-1}\underline{\iota}+\frac{1}{k+1}\mathbf{\Lambda}^{-1}(\mathbf{I}-(\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}_i^{-1})\underline{v}_i+ \\ & \frac{1}{2}\mathbf{\Lambda}^{-1}(\mathbf{I}-(\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}_i^{-1})(\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}_i^{-1}\frac{k-1}{k+1}\underline{v}.\end{aligned} \tag{IA.15}$$

The last term equals zero and hence, we have

$$\underline{o}_i = Z_i(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}\mathbf{\Lambda}^{-1}\underline{\iota}+\frac{1}{k+1}\mathbf{\Lambda}^{-1}(\mathbf{I}-(\underline{\iota}'\mathbf{\Lambda}_i^{-1}\underline{\iota})^{-1}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}_i^{-1})\underline{v}_i. \tag{IA.16}$$

□

The result in Lemma 4 is intuitive. With multiple informed investors, each investor anticipates the other investors' order flow. Because other investors' order flow contributes to price impact, each investor scales back the speculative component of order flow. Moreover, the funding component for each investor is now smaller because each investor scales back speculative order flow. We can pre-multiply with $\mathbf{\Lambda}$, solve for $o_{i,1}$, re-work as before, and aggregate to the market level to get the same result as in Proposition 1, up to a scalar multiplication:

Proposition 2. *The average order flow-implied private information shock vector* $\underline{\bar{v}}$ *with* k *perfectly informed investors is given by:*

$$\underline{\bar{v}} = \frac{k+1}{k}\mathbf{\Lambda}\underline{o} - \frac{k+1}{k}\lambda_1 o_1 \underline{\iota}, \tag{IA.17}$$

where $\underline{\bar{v}}$ *is the vector containing the average signal among all informed investors.*

Proof. Pre-multiplying $\underline{o}_i$ by $\mathbf{\Lambda}$ yields

$$\mathbf{\Lambda}\underline{o}_i = \frac{1}{k+1}\underline{v}_i + (\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}(Z_i - \underline{\iota}'\mathbf{\Lambda}^{-1}\underline{v}_i)\underline{\iota} \Rightarrow \tag{IA.18}$$

$$\underline{v}_i = (k+1)\mathbf{\Lambda}\underline{o}_i - (k+1)\lambda_1 o_{i,1}\underline{\iota}. \tag{IA.19}$$

Aggregating over the market yields

$$k\underline{\bar{v}} = (k+1)\mathbf{\Lambda}\underline{o} - (k+1)\lambda_1 o_1 \underline{\iota} \Rightarrow \tag{IA.20}$$

$$\underline{\bar{v}} = \frac{k+1}{k}\mathbf{\Lambda}\underline{o} - \frac{k+1}{k}\lambda_1 o_1 \underline{\iota}. \tag{IA.21}$$

□

Hence, as long as k is known to the econometrician, $\underline{\bar{v}}$ can be perfectly reverse-engineered. If k is not known (but common across stocks), one can still rank stocks with respect to private information based on $\mathbf{\Lambda}\underline{o}$. Note that our baseline model is a special case of this extension with $k = 1$.

Internet Appendix IA.2.3. Noisy common signal with multiple informed investors

One can also introduce noise into the setting with multiple investors. This can be done in multiple ways, depending on the assumptions made on the correlation structure of the noise. The setting with perfectly correlated noise across informed investors is easiest to work out and is presented here. As before, if the number of informed investors k is known, one can reverse-engineer the average (discounted) signal.

For the setting with multiple informed investors and a common noisy signal, we define ϕ_j as in Eq. (IA.2), to construct $\tilde{v}_{i,j}$ for all investors i with a signal. Since $\tilde{v}_{i,j} = \tilde{v}_j$ for all informed investors i, we can simply use the results from Internet Appendix IA.2.2, but with $\underline{v}_i$ replaced by $\underline{\tilde{v}}_i$:

Corollary 2. *In the presence of signal noise and k informed investors per security with a common signal, the best estimate of order flow-implied private information is given by:*

$$\underline{\tilde{v}} = \frac{k+1}{k}\mathbf{\Lambda}\underline{o} - \frac{k+1}{k}\lambda_1 o_1 \underline{\iota}. \tag{IA.22}$$

Proof. We have that $E(\underline{\delta}|\underline{v}_i) = \underline{\tilde{v}}_i$. Hence, the Lagrangian is equal to Eq. (IA.5), but with $a_i^{mult} = \underline{\tilde{v}}_i$. It immediately follows from Lemma 4 and Proposition 2 that Eq. (IA.22) must hold. □

Hence, the result of Internet Appendix IA.2.2 directly carries over, with the only difference that the discounted rather than the actual signal can be reverse-engineered.

Internet Appendix IA.2.4. Noisy heterogeneous signal with multiple informed investors

Now assume that the noise in a signal about security j is independently and identically distributed across informed investors. As before, each investor now scales her signal to obtain the conditional expected value of the information in the signal:

$$E_i(\delta_j|v_{i,j}) = \phi_{i,j} v_{i,j} = \tilde{v}_{i,j}. \tag{IA.23}$$

Because noise is identically distributed across informed investors, $\phi_{i,j} = \phi_j = \frac{\sigma^2_{F_j}}{\sigma^2_{F_j}+\sigma^2_{u_j}}$. However, because noise is now independently distributed, we obtain:

$$E(v_{l,j}|v_{i,j}) = \phi_j v_{i,j}, \Rightarrow \tag{IA.24}$$

$$E(\tilde{v}_{l,j}|v_{i,j}) = \phi_j^2 v_{i,j} = \phi_j \tilde{v}_{i,j}. \tag{IA.25}$$

for any informed investor $l \neq i$. We can substitute this expression and solve for optimal order flow $\underline{o}_i$:

Lemma 5. *In the presence of k heterogeneously informed investors, optimal order flow for informed investor i is given by*

$$\underline{o}_i = Z_i \mathbf{\Lambda}^{-1} \underline{\iota}(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1} + diag(2\underline{\iota} + \underline{\phi}(k-1))^{-1}\mathbf{\Lambda}^{-1}(\mathbf{I} - (\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}\underline{\iota}\underline{\iota}'\mathbf{\Lambda}^{-1})\underline{v}_i, \tag{IA.26}$$

where $diag(\cdot)$ is a diagonal matrix with the elements of the argument vector on the diagonal.

Proof. Up to Eq. (IA.10) included, the proof is identical to that of Lemma 4. We have that

$$E(v_{l,j}|v_{i,j}) = \phi_j v_{i,j}, \Rightarrow \tag{IA.27}$$

$$E(\tilde{v}_{l,j}|v_{i,j}) = \phi_j^2 v_{i,j} = \phi_j \tilde{v}_{i,j}. \tag{IA.28}$$

for any informed investor $l \neq i$. Expected order flow by investor l is purely driven by speculative and funding flow, which in turn is proportional to $\underline{\tilde{v}}_l$. Since $E(\underline{\tilde{v}}_l|\underline{\tilde{v}}_i)$ is now shrunk by a factor $\underline{\phi}$, this should also hold for $E(\underline{o}_{-i})$. Substituting and working through as in Lemma 4 yields Eq. (IA.26). $\square$

We can pre-multiply with $\mathbf{\Lambda}$, solve for $o_{i,1}$, re-work as before, and aggregate to the market level to get the same result as in Proposition 1, up to a scalar multiplication:

Proposition 3. *In the presence of signal noise and k informed investors per security with heterogeneous signals, we have that the best estimate of order flow-implied private information is given by:*

$$\bar{\underline{\tilde{v}}} = \frac{1}{k} diag(2\underline{\iota}+\underline{\phi}(k-1))(\mathbf{\Lambda}\underline{o}-\lambda_1 o_1 \underline{\iota}). \tag{IA.29}$$

Proof. Pre-multiplying $\underline{o}_i$ by $\mathbf{\Lambda}$ yields

$$\mathbf{\Lambda}\underline{o}_i = diag(2\underline{\iota}+\underline{\phi}(k-1))^{-1}\underline{v}_i+(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}(Z_i-\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{v}_i)\underline{\iota} \Rightarrow \tag{IA.30}$$

$$\underline{v}_i = diag(2\underline{\iota}+\underline{\phi}(k-1))^{-1}(\mathbf{\Lambda}\underline{o}_i-\lambda_1 o_{i,1}\underline{\iota}). \tag{IA.31}$$

Aggregating over the market yields

$$k\bar{\underline{v}} = diag(2\underline{\iota}+\underline{\phi}(k-1))^{-1}(\mathbf{\Lambda}\underline{o}-\lambda_1 o_1 \underline{\iota}). \tag{IA.32}$$

$\square$

Hence, as long as $\underline{\phi}$ and k are known by the econometrician, average private information shocks can be perfectly reverse-engineered. Even if $\underline{\phi}$ and k are not known, one can rank stocks on their private information based on $\mathbf{\Lambda}\underline{o}$ if $\underline{\phi} = \phi\underline{\iota}$ (i.e., the signal to noise ratio is the same for all stocks).

Internet Appendix IA.2.5. Discussion on separation of roles and costless credit lines

The results in Lemma 2 on the optimal order flow of investor i in the baseline model are intriguing. The magnitude of the liquidity-motivated order flow component does not depend on the information shock and vice versa. Hence, noise traders and informed traders can be perfectly separated as in Kyle (1985). The reason we also exposed informed investors to liquidity shocks was (i) because funding flow generates a liquidity shock like order flow component and (ii) to highlight the generality of our results. We can even shut down the funding need for informed investors (assuming they have a cost-free credit line) to get very close to a multivariate Kyle (1985) setting (the main difference being strategic noise traders). As Internet Appendix IA.3 shows, it would be even easier to attain a market equilibrium with linear price impact. Finally, one could also allow a certain fraction of investors (but not all) to have access to a cost-free credit line. Uninformed order flow in that case would still be sufficiently summarized by $\lambda_1 o_1$.

Internet Appendix IA.3. Market equilibrium and linear price impact

In this Appendix, we extend our baseline model in Section 2 by deriving the equilibrium price impact parameters for all securities from a system of equations. It is not trivial that a solution exists to such a system of equations. If noise traders become price-sensitive in a traditional Kyle (1985) setting, altering price impact becomes a less effective tool for market makers to manage their adverse selection risk as increasing price impact reduces the profits from providing liquidity to uninformed order flow in exactly the same proportion as losses from informed flow are avoided. As a result, a market equilibrium with linear price impact only materializes when the ratio of signal variance to the sum of signal variance and noise trader variance equals exactly one half. The literature has come up with several ways to address this issue. In Kyle and Obizhaeva (2018), information production is endogenized such that, in equilibrium, the ratio of signal variance to the sum of signal variance and noise trader variance equals one half (in other words, signal variance equals noise trader variance). In Spiegel and Subrahmanyam (1992) and Lee and Kyle (2018), price-sensitive noise trading demand stems from a desire to hedge endowment risk (rather than from liquidity-motivated and funding-induced trading as in our model). In their models, noise trading and speculative trading are affected by price impact in different ways. As a result, a market equilibrium can still be found under the condition that the ratio of signal variance to noise trader variance is not too high.

In our model, price impact affects demand in the usual way (inversely proportional to the price impact parameter), but also in two other ways. A higher price impact lowers the total absorption capacity for uninformed order flow because of which each security needs to absorb more price impact (this is the $(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}$ term). This partially counters the reduction of uninformed flow as a result of increased price impact and thereby may allow for an equilibrium to exist. Finally, there is the funding flow that is also inversely proportional to price impact. As a result, the inclusion of funding flow makes it harder to find an equilibrium.

There is another issue in our setup, which is that for internal consistency reasons, there should be another source of price impact besides adverse selection. The reason is that the information-free benchmark security by definition is not exposed to adverse selection, but still needs to have a strictly positive price impact (otherwise all uninformed flow would be allocated there). We achieve that by combining inventory risk as in Stoll (1978) with adverse selection risk. This requires market makers to be risk-averse. In particular, we set up a model along the lines of Madhavan and Smidt (1991a), where we restrict fixed transaction costs (the bid-ask spread) to zero, in line with our basic model setup. Price impact is then the sum of price impact due to asymmetric information (as in Kyle (1985)) and price impact due to inventory and risk aversion (as in Stoll (1978)).

As in our baseline model, we assume that, for every security j, there is exactly one investor that receives a perfectly informative signal $v_j \sim N(0, \sigma^2_{v_j})$. For the liquidity shocks, we now make the additional distributional assumption that liquidity shocks Z_i are distributed according to $N(0, \sigma^2_Z)$ and that liquidity shocks are uncorrelated across the M investors, such that $\sum_{i=1}^{M} Z_i$ is distributed according to $N(0, M\sigma^2_Z)$. As in the baseline model, we assume that all information shocks are orthogonal to each other and to all liquidity shocks.[IA.3]

Under these assumptions, we can derive the best estimate of the value of a security $\tilde{v}_j$ by the market maker, conditional on observing order flow $\tilde{o}_j = o_j$:

$$\begin{aligned} E(\tilde{v}_j|o_j) &= E(\tilde{v}_j) + \frac{cov(\tilde{v}_j, \tilde{o}_j)}{var(\tilde{o}_j)}(\tilde{o}_j - E(\tilde{o}_j)), \\ &= \bar{v} + \frac{\beta_j \sigma^2_{v_j}}{\beta^2_j \sigma^2_{v_j} + \sigma^2_{uninf_j}} \tilde{o}_j, \end{aligned} \tag{IA.1}$$

in which, by projection, β_j denotes the demand elasticity of information shocks for security j and is given by $\beta_j = \frac{1}{2\lambda_j}$ in equilibrium as derived in the basic model setup. Working out $\sigma^2_{uninf_j}$, we obtain:

$$\sigma^2_{uninf_j} = \frac{1}{\lambda^2_j}(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-2}(M\sigma^2_Z + \frac{1}{4}\sum_k \lambda_k^{-2}\sigma^2_{v_k}), \tag{IA.2}$$

$$= \beta^2_j(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-2}(4M\sigma^2_Z + \sum_k \lambda_k^{-2}\sigma^2_{v_k}). \tag{IA.3}$$

Price impact only due to market maker inventory concerns as in Stoll (1978) is given by $\frac{A}{2}\sigma^2_j$, where A denotes market maker risk aversion, and σ^2_j denotes return variance of security j. We need such inventory motivated price impact in order to ensure that the information-free security also has a strictly positive price impact. We add (purely)

[IA.3] Allowing for non-zero correlations among these shocks is possible, but comes at the expense of additional complexity.

inventory motivated price impact as in Stoll (1978) and (purely) adverse selection motivated price impact as derived in Eq. (IA.1) (that is, the term in front of $\tilde{o}_j$) to obtain the following system of equations:

$$\begin{aligned}\lambda_j &= \frac{A}{2}\sigma_j^2 + \frac{\beta_j \sigma_{v_j}^2}{\beta_j^2(\sigma_{v_j}^2 + (\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-2}(4M\sigma_Z^2 + \sum_k \lambda_k^{-2}\sigma_{v_k}^2))}, \\ &= \frac{A}{2}\sigma_j^2 + \frac{\sigma_{v_j}^2}{\beta_j(\sigma_{v_j}^2 + (\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-2}(4M\sigma_Z^2 + \sum_k \lambda_k^{-2}\sigma_{v_k}^2))}, \end{aligned} \tag{IA.4}$$

for all securities j. The adverse selection related term is similar to that in Kyle (1985), but with a twist. In the regular Kyle (1985) model, only speculative order flow is sensitive to price impact, because of which only the signal size variance $\sigma_{v_j}^2$ is multiplied with β_j^2. However, in our setting, also uninformed order flow is sensitive to price impact and, hence, the whole denominator of the second term is a factor of β_j^2. Substituting $\beta_j = \frac{1}{2\lambda_j}$ and rewriting yields:

$$\frac{A}{2}\sigma_j^2 = \lambda_j \left(1 - \frac{2\sigma_{v_j}^2}{\sigma_{v_j}^2 + (\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-2}(4M\sigma_Z^2 + \sum_k \lambda_k^{-2}\sigma_{v_k}^2)}\right) \tag{IA.5}$$

$$\frac{A}{2}\sigma_j^2 = \lambda_j \left(\frac{(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-2}(4M\sigma_Z^2 + \sum_k \lambda_k^{-2}\sigma_{v_k}^2) - \sigma_{v_j}^2}{\sigma_{v_j}^2 + (\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-2}(4M\sigma_Z^2 + \sum_k \lambda_k^{-2}\sigma_{v_k}^2)}\right) \Rightarrow \tag{IA.6}$$

$$\lambda_j = \frac{A}{2}\sigma_j^2 \left(\frac{\sigma_{v_j}^2 + (\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-2}(4M\sigma_Z^2 + \sum_k \lambda_k^{-2}\sigma_{v_k}^2)}{(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-2}(4M\sigma_Z^2 + \sum_k \lambda_k^{-2}\sigma_{v_k}^2) - \sigma_{v_j}^2}\right), \forall j. \tag{IA.7}$$

We note that (i) the cross-section of price impact parameters must be determined simultaneously in equilibrium as these affect each other (which is not the case in Kyle, 1985), and (ii) asymmetric information and inventory channels also show up multiplicatively in determining price impact (whereas such effects are only additive in e.g., Madhavan and Smidt (1991b)). The final expression in Eq. (IA.7) is hard to solve analytically and does not have a guaranteed solution. Yet, we can solve the system of equations numerically and show that there are cases in which a solution exists.

Proposition 4. *There are ranges for parameter values σ_{v_j}, σ_j, σ_Z, m, and A for which there exists a $\mathbf{\Lambda}$ that solves the system of equations stipulated in (IA.7).*

Proof. The existence for one set of parameters is sufficient to prove the proposition. We set parameters $\sigma_{v_j} = 0.15 \forall j \neq 1$, $\sigma_{v_1} = 0$, $\sigma_j = 0.25 \forall j$, $\sigma_Z = 0.25$, $m = 1,000,000$, $n = 70$, and $A = 1$. Solving Eq. (IA.7) for $j = 1$ yields $\lambda_1 = \frac{A}{2}\sigma^2 = 0.0313$. We fix this value and then use the **fsolve** function in Matlab to solve the remaining $n-1$ equations. This yields that $\lambda_j = 0.0470 \; \forall j \neq 1$. □

As said, the system of equations in (IA.7) is not guaranteed to have a solution. If noise trading and informed trading would both scale with the inverse of price impact in the same way, no solution would exist. Yet, the liquidity-induced noise trading (resulting from σ_Z) needs to be allocated somewhere. Now as λ_j increases, $(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}$ also increases because it contains λ_j. This offsetting effect is sufficient to allow for solutions in some parameter ranges. The funding flows are also affected by λ_j in two offsetting ways. First, it also contains the term $(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}$, and thereby helps for the existence of a solution. However, the size of the funding flow is declining in λ_j, which generates the opposite effect. On net, the latter effect dominates, but the net effect is small.

In general, solutions are more likely to exist when

1. the number of assets N is small
2. signal variances $\sigma_{v_j} \forall j \neq 1$ are close
3. the number of investors M is large
4. liquidity shock volatility σ_Z is large

All these effects are substitutes in the sense that there is a wider range of signal variances that provide a solution when N is small than when N is large. Similarly, there is a wider range of signal variances that provide a solution when M and/or σ_Z are large than when those are small. The intuition for these comparative statics is the following. In order to find a solution, there should be sufficient noise trading volume in an asset. Expected noise trading volume increases in M and σ_Z. The reason that we can find a solution comes from the impact of λ_j on $(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}$. When λ_j is large compared to other stocks, its impact on $(\underline{\iota}'\mathbf{\Lambda}^{-1}\underline{\iota})^{-1}$ is smaller. Similarly, when there are more stocks (higher N), the impact of λ_j on $(\iota'\mathbf{\Lambda}^{1}\iota)^{-1}$ is smaller (everything else equal).

Concluding, this Appendix shows that in a setup with strategic uninformed trading in the cross-section of securities and competing market makers, market equilibria can materialize with pricing rules that involve linear price impact, but no fixed transaction costs in the form of a bid-ask spread. This confirms the internal consistence of our results.