November 30, 2014

Cross-sectional identification of informed trading

Dion Bongaerts, Dominik Rösch, and Mathijs van Dijk*

Abstract

We propose a new approach to measuring informed trading in individual securities based on a portfolio optimization model for investors facing information and liquidity shocks. These shocks induce speculative and liquidity-motivated order flow, taking into account the price impact of trading. The model allows us to back out the amount of informed trading from a security's aggregate order flow, based on the cross-section of price impact parameters (λ) and order imbalances (OIB). Furthermore, we obtain a very simple expression for a security's aggregate private information shock: its $\lambda \times OIB$, in excess of the same term for a benchmark security that is insulated from informed trading. We validate our private information measure (based on daily data for all S&P 1500 stocks over 2001-2010) by showing that it is strongly related to contemporaneous returns, and that return reversals are significantly weaker following stock-days with high private information estimates.

*Bongaerts, Rösch, and van Dijk are at the Rotterdam School of Management, Erasmus University. E-mail addresses: dbongaerts@rsm.nl, drosch@rsm.nl, and madijk@rsm.nl. 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 firms whose securities are more subject to informed trading face greater illiquidity in these securities' secondary markets, a higher cost of capital, and reduced incentives to invest.[1] However, measuring private information and informed trading empirically remains a considerable challenge.

In this paper, we propose a new way of measuring informed trading based on a portfolio optimization model for individual investors. Our approach has two main advantages. First, it allows us to identify the amount of informed trading in an individual security over a given period based on the cross-section of price impact parameters (λ) and order imbalances (OIB, or the volume of buyer- minus seller-initiated trades). Hence, our measure can be estimated for each security on each day, or even at higher frequencies. Second, our model also delivers a very simple and intuitive expression for the aggregate private information shock for a given security over a given period. In other words, in addition to estimating the prevalence of trading based on private information, we can measure the direction and magnitude of private information for each security on each day.

In the model, investors arrive at the market with an optimal portfolio of securities, but are then hit by liquidity shocks and private information shocks that induce them to rebalance their portfolio. Investors' order flow generates price impact that is linear in trading volume, which implies that total transaction costs are quadratic in trading volume. Individual securities differ in their price impact parameter for exogenous reasons. When hit by a liquidity shock, investors optimally spread their trading over many securities, such that the marginal transaction costs for all securities are equal. As a result, the order flow in individual securities is proportional to the inverse of their price impact parameter, which implies that most trading is done in the most liquid securities. When

[1] See, among many others, 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 Gumbel (2008), and Edmans (2009).

hit by a private information shock about a certain security, investors trade an amount in that security that is inversely related to its price impact parameter. Furthermore, investors trade other securities in the opposite direction to finance the speculative trade, where again the amount of trading in each security is inversely related to its price impact.

The aggregate order flow (across all investors) in a security thus consists of three components: (i) liquidity-motivated order flow, (ii) speculative order flow based on private information about that security, and (iii) "funding" order flow to finance the speculative trading in other securities. When we introduce a benchmark security that is insulated from informed trading to resolve underidentification, we obtain a closed-form solution to back out the amount of informed trading in any security, or component (ii), from its aggregate order flow and the aggregate order flows and price impact parameters of other securities.

We refer to our identification of informed trading as "cross-sectional" since it exploits the idea that order flow that is purely liquidity-motivated has the same sign for all securities, while trading based on private information about a certain security results in opposite-sign order flow in other securities to finance the speculative trade. Crucially, the identification of informed trading also makes use of the notion that any order flow is affected by the expected price impact of trading.

Empirically, our model allows us to measure the dollar volume of informed trading in any security over any time period based on the cross-section of price impact parameters and order imbalances for a relevant set of peer securities as well as a benchmark security. We can also compute the probability of informed trading inferred from the cross-section (or $XPIN$) as the fraction of informed trading over total trading.

Next to a measure of the volume (and probability) of informed trading, our model also provides a very simple expression for a security's aggregate private information shock (aggregated across investors): the security's order imbalance multiplied by its price impact parameter ($\lambda \times OIB$), minus the order imbalance of the benchmark security multiplied by the benchmark's price impact parameter. The intuition is that the observed order imbalance in a security is more likely to be information-driven when the price impact of trading is high, since investors only trade securities that are expensive to trade

when they have valuable private information about these securities. Furthermore, any trading in the benchmark security is either liquidity-motivated or funding-motivated, so the benchmark's order imbalance (accounting for its price impact) forms a natural reference point that can be used to isolate the aggregate private information shock of an individual security.

We estimate our measures of the amount and probability of informed trading and of the aggregate private information shock for all S&P 1500 stocks each day in the period 2001-2010 based on intraday price and transaction data from the NYSE Trade and Quote (TAQ) database. We estimate daily price impact parameters based on intraday data by implementing the approach of Glosten and Harris (1988). We use each stock's moving average price impact estimate over the past 20 days as the expected price impact on the current day. We estimate the daily order imbalances of individual stocks by signing individual trades using the Lee and Ready (1991) algorithm. Our final sample consists of all 2,130 stocks (listed at NYSE, Nasdaq, or Amex) that were an S&P 1500 constituent at some point during our sample period of 2001-2010 and that survive our basic data screens. As the benchmark security, we use the SPDR S&P500 ETF (ticker "SPY"), for which we obtain consolidated trades and quotes from the Thomson Reuters Tick History (TRTH) database. 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 (Baker and Stein, 2004), and since the SPDR is unlikely to be used for trading on private information of individual securities.[2]

The main purpose of our empirical analyses is to assess whether cross-sectional patterns in stock returns are consistent with our private information measure picking up meaningful cross-sectional variation in aggregate private information shocks. As our key predictions are cross-sectional in nature, most of our tests are based on a further simplified version of our private information measure: a stock's order imbalance multiplied by its price impact parameter ($\lambda \times OIB$). Since the correction for the benchmark's order im-

[2]This idea is similar to the rationale behind program trading facilities. These also allow better liquidity because at least 15 securities need to be traded at the same time and hence the likelihood of trading on private information on any of these securities is low.

balance times its price impact is the same for all stocks on a given day, this simplification does not affect our cross-sectional tests.

We first show, in Fama-MacBeth regressions, that the cross-section of daily stock returns is positively and highly significantly related to this simplified private information measure for individual stocks estimated on the same day. This finding is consistent with the idea that stocks with a more positive (negative) information shock on a given day have a more positive (negative) realized stock return, but it does not rule out other interpretations of our private information measure. In particular, our measure is a positive function of a stock's order imbalance and it is well-known that stocks with a more positive (negative) order imbalance on a given day tend to have a more positive (negative) return, for reasons that may be distinct from private information (e.g., price pressure). However, we show that the positive relation between the cross-section of stock returns and our private information measure survives controlling for order imbalance and expected price impact separately. In other words, $\lambda \times OIB$ has explanatory power for the cross-section of returns that goes beyond that of λ and OIB individually. We are not aware of models that provide an alternative interpretation of $\lambda \times OIB$. Furthermore, the explanatory power of $\lambda \times OIB$ is not subsumed by other "scaled" measures of order imbalance, such as the product of OIB and the quoted bid-ask spread or OIB scaled by market capitalization.

We then follow the reasoning that if our measure picks up private information, return reversals should be weaker following stock-day observations for which our measure assumes large negative or large positive values. After all, the price impact of informed trades should be permanent, while the price impact of uninformed order flow should be temporary (e.g., Kyle, 1985; Admati and Pfleiderer, 1988; Glosten and Harris, 1988; Sadka, 2006). To test this conjecture, we run daily Fama-MacBeth regressions of the cross-section of stock returns on one-day lagged returns, interacted with the absolute value of $\lambda \times OIB$. We reproduce the common result in the literature that the one-day autocorrelation in returns is negative (e.g., Roll, 1984; Cox and Peterson, 1994; Nagel, 2012). The interaction effect between one-day lagged returns and the absolute value of $\lambda \times OIB$ is significantly positive, indicating that returns revert significantly less following

stock-days with large negative or large positive values of the private information measure.

To get a better idea of the economic magnitude of the reduced return reversal for high private information shocks, we also take a double-sorting approach to studying the relation between return reversals and the private information measure. We first sort stocks into quintile portfolios based on their private information measure $\lambda \times OIB$ on a given day, in such a way that portfolio 1 and 5 contain stocks with, respectively, large negative and large positive values for the measure. We then sorts stocks within each quintile into winner and loser stocks based on their returns on that day. We compute the daily returns on a reversal strategy within each quintile portfolio based on a long position in that day's loser stocks and a short position in that day's winner stocks, held from the market close on that day till the market close on the next day. The results of this double sort show that the abnormal returns (alphas) on the reversal strategy of quintile portfolio 3 (consisting of stocks with values of the private information measure close to zero) are significantly greater than the abnormal returns on the reversal strategy in the two extreme private information portfolios (quintiles 1 and 5). The economic magnitude of the difference in the strength of the return reversals is substantial, at 12 basis points per day. We interpret this as further evidence consistent with the view that our measure picks up meaningful cross-sectional variation in the direction and magnitude of private information for individual stocks.

In sum, this paper proposes new measures for the amount and probability of informed trading in individual stocks based on a portfolio optimization model whose key predictions concern the cross-section of order imbalances and price impact parameters. The model also yields a simple measure of the direction and magnitude of private information for individual stocks. We provide empirical support for this measure by showing that it is positively related to contemporaneous stock returns in the cross-section, and that return reversals are significantly weaker following stock-days with high values for this measure.

We contribute to the literature on measuring informed trading by suggesting an alternative to the popular "probability of information-based trading" (PIN) measure developed by Easley, Kiefer, O'Hara, and Paperman (1996) and Easley, Hvidkjaer, and O'Hara (2002), which is based on a market microstructure model instead of a portfolio

optimization model and which has a different intuition. An advantage of our approach to measuring informed trading is that it is easy to implement and that it does not require a long time-series of transaction data for individual securities (and can thus be estimated even at high frequencies), since its main data requirements are of a cross-sectional rather than a time-series nature. Our work is also related to more recent papers on the "volume-synchronized probability of informed trading" or $VPIN$, see, among others, Easley, Lopez de Prado, and O'Hara (2011, 2012). A common feature of $VPIN$ and our measure of information trading is that order imbalances play a key role, but our measure is distinct in that it also takes into account the price impact of trading.

Furthermore, to the best of our knowledge, our study is the first to propose a way to measure the magnitude and direction of the aggregate private information shock in an individual security contained in its trading over a given period. Our approach complements the work of, among others, Glosten and Harris (1988), Hasbrouck (1991), and Sadka (2006), who measure the information effects of a trade through its permanent price impact, but who do not attempt to extract a direct proxy for the private information shocks on which informed trades are based.

The paucity of sophisticated proxies for informed trading and private information is illustrated by the paper of Lai, Ng, and Zhang (2014), who benchmark PIN using crude, low-frequency firm-level proxies for information asymmetry such as the number of analysts following the firm, the analyst forecast dispersion, the age of the firm, and equity index membership. We hope that our new, high-frequency measures of informed trading and private information provide useful alternatives to existing measures and offer new opportunities to test and revise existing private information models of market microstructure, asset pricing, and corporate finance.

2. Basic model assumptions and notation

In this section, we introduce the basic setup for the theoretical portfolio optimization model from which we deduce the market implied information per security to be incorporated in prices.

Our model covers one period and concerns a market for N securities. These securities

are typically risky, but a riskless security can be included. The returns on the securities are collected in the vector $\mathbf{r}$ and follow a multivariate lognormal distribution with means and covariance matrix $E(\mathbf{r})$ and $\mathbf{\Sigma}$, respectively. Let us for notational convenience define $\boldsymbol{\sigma^2}$ as the array that contains the diagonal elements of $\mathbf{\Sigma}$.

There are M investors in the market, which are indexed with i. Each investor i has power utility with CRRA parameter γ_i and starting wealth W_i. We assume that investors arrive to the market with an optimal starting portfolio. Moreover, we assume that investors cannot dislocate their portfolio so much that individual securities start to dominate portfolio such that idiosyncratic risk is beyond concern . Investors are exposed to liquidity shocks as well as potential private information shocks. Liquidity shocks Z_i arrive randomly and are expressed as a fraction of initial wealth W_i such that $Z_i > 0$ corresponds to money inflow. If no shock arrives, $Z_i = 0$. Information shocks are described in more detail below. Given the liquidity and information shocks, each investor i has to determine optimal holdings $\mathbf{x_i}$ of all securities. His starting portfolio allocation is denoted by $\mathbf{x_i^*}$.

Trading demands of investors are accommodated by a financial intermediation sector (i.e., market makers) for a fee. In particular, order flow $o_{i,j}$ of investor i in security j has price impact on security j when it is traded. This leads to a lower expected return (without affecting risk), which increases linearly with trade size. More explicitly, we express total price impact $\psi_j(o_{i,j})$ as:

$$\psi_j(o_{i,j}) = \lambda_j \delta_{i,j} o_{i,j} \; \forall \; j, \tag{1}$$

where λ_j is the price impact parameter for security j expressed in percentage points lower expected return over the average investor horizon per dollar traded and $\delta_{i,j}$ is a trade sign indicator for the trade by investor i in security j. Total trading costs are then given by multiplying the average shortfall or excess in price with the size of the transaction:

$$|o_{i,j}\psi_j(o_{i,j})| = \delta_{i,j} o_{i,j} \lambda_j \delta_{i,j} o_{i,j} \; \forall \; j. \tag{2}$$

Hence, total transaction costs (execution shortfall) are quadratic in order flow sent by

an investor. We define the matrix $\mathbf{\Delta_i}$ as a diagonal matrix with $\delta_{i,j}$ as its jth diagonal element. Similarly, we define $\mathbf{\Lambda}$ as the matrix that contains the λ_js on its diagonal. We assume that for all j, $\lambda_j > 0$, also for the riskless security (if any).

3. Individual investor portfolio optimization

3.1. Liquidity shocks only

We take a somewhat unconventional approach to portfolio optimization. We assume a CAPM-like setting in which investors may be heterogeneous (due to for example background risk) and have an optimal portfolio allocation x_i^*, given information at time 0. Moreover, we assume that all securities are correctly priced; thus, $(E(r-r_f)+\frac{1}{2}\sigma^2)/\beta = \iota(E(r_m = r_f)+\frac{1}{2}\sigma_m^2) = \iota\zeta$, where ζ is the market risk premium. Under these assumptions, we can let investors optimize risk-adjusted portfolio returns.[3] When we do this, we need to impose a budget constraint to avoid that the investor loads up on risk. Combined with transaction costs, the investor would like to keep his portfolio as it is. Our motivation to use a static model with somewhat incomplete preferences is that this will give very neat and tractable solutions under relatively mild assumptions.

An investor only receiving a liquidity shock Z_i optimizes:

$$\max_{\mathbf{x_i}} \mathbf{x_i}'\iota\zeta - \frac{1}{1+Z_i}(W_i(\mathbf{x_i}(1+Z_i)-\mathbf{x_i^*})'\mathbf{\Delta_i}\mathbf{\Lambda}\mathbf{\Delta_i}(\mathbf{x_i}(1+Z_i)-\mathbf{x_i^*})), \tag{3}$$

subject to the budget constraint

$$\iota'\mathbf{x_i} = 1. \tag{4}$$

We note that this way of formulating the rebalancing decision problem is intuitive and

[3]This approach differs from the traditional mean-variance portfolio optimization problem in that the covariance matrix is not explicitly taken into account. As such, it looks a bit like a risk-neutral setting, except for the fact that we make risk-adjustments by standardizing by β. Our motivation for doing this is to keep the model tractable and to avoid instability due to estimation error of individual elements of $\mathbf{\Sigma}$. Otherwise, in solving for optimal portfolio weights, we need to invert an investor-specific weighted sum of $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$, which is highly non-linear and complex. The downside of this approach is that investors could end up with concentrated portfolios since additional diversification is not rewarded (but complete diversification is assumed). However, systematic risk is taken into account since $E(\mathbf{r})$ is scaled by β.

parsimonious. As $\mathbf{\Lambda}$ and $\mathbf{\Delta_i}$ are diagonal matrices, their order of multiplication in (3) can be changed. As a result, since $\mathbf{\Delta_i \Delta_i} = \mathbf{I}$, the "endogenous" parameter matrix $\mathbf{\Delta_i}$ drops out from the price impact part and we obtain a solution without any endogenous parameters.[4] Another way of seeing this is that price impact is linear in signed order flow, such that total transaction costs are quadratic in signed order flow, so that taking absolute values is irrelevant.

The problem can be optimized by standard constrained optimization techniques involving a Lagrangian multiplier.[5] The optimal portfolio weights are given by the following Lemma:

Lemma 1. *The solution to optimization problem (3) is given by:*

$$\mathbf{x_i} = \mathbf{Q_i}^{-1}\iota\zeta + \mathbf{Q_i}^{-1}2W_i\mathbf{\Lambda}\mathbf{x_i^*} - \mathbf{Q_i}^{-1}\iota\zeta + \mathbf{Q_i}^{-1}\iota f\frac{Z_i}{1+Z_i} \tag{5}$$

$$= \frac{1}{1+Z_i}\mathbf{x_i^*} + \frac{Z_i}{1+Z_i}\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}. \tag{6}$$

Proof. See Appendix. □

The new portfolio holdings are therefore equal to the portfolio holdings in case the liquidity shocks could be settled with a risk and friction free savings account (first term) plus a transaction cost driven adjustment (second term). This second term consists of the relative size of the shock $\left(\frac{z_i}{1+Z_i}\right)$ times the fraction of the shock that is accommodated by every security. This fraction always lies between 0 and 1 and is proportional to the inverse of the price-impact of the security, such that most trading is done in the most liquid securities.

[4] Bongaerts, De Jong, and Driessen (2011) use a similar setting, but their model still features these endogenous parameters since they focus on bid-ask spreads rather than on price impact.

[5] Note that incorporating other constraints, such as short sale constraints, in this framework is convenient, but comes at the cost of increased complexity. The Lagrangian multiplier μ in the proof can be interpreted as a shadow price. In this case, it is the utility loss to the investor in optimal solutions compared to the setting in which shocks can also be accommodated with a transaction cost-free risk-free account.

Individual order flow is now given by:

$$\mathbf{o_i} = W_i(1+Z_i)\mathbf{x_i} - W_i\mathbf{x_i^*} \tag{7}$$

$$= W_i Z_i \mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}. \tag{8}$$

One can verify that this is indeed the optimal order flow. If we pre-multiply (8) by $\mathbf{\Lambda}$, we see that the solution yields order flows such that the marginal transaction costs for all securities are equal, as the RHS consists solely of scalars multiplied with a unity vector. Thus, it is impossible to sell a bit more of one security and a bit less of another and thereby be better off.

3.2. Adding information shocks

We now introduce an information shock that will create a Jensen's alpha (standardized by β) of $\mathbf{v_i}$ on the securities. In other words, in addition to the liquidity shock, each investor i receives an information shock $\mathbf{v_i}$, which is essentially a vector of the alphas gross of transaction costs that can be generated for each security. The solution to the investor optimization problem is then given by the following Lemma.

Lemma 2. *With liquidity and private information shocks, optimal portfolio weights are given by*

$$\mathbf{x_i} = \frac{1}{1+Z_i}\mathbf{x_i^*} + \frac{Z_i}{1+Z_i}\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1} + \frac{1}{2W_i(1+Z_i)}\mathbf{\Lambda}^{-1}(\mathbf{I} - (\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})\mathbf{v_i}. \tag{9}$$

Proof. See Appendix. □

It is worthwhile analyzing the various components of this solution. The first two components are identical to the case with only liquidity shocks. The third term consists of three parts. The first part is $\mathbf{\Lambda}^{-1}\mathbf{v_i}$. This is the solution to $\Lambda y_i = v_i$, which is a first order optimality condition as it equates for each security marginal benefits (alpha return) of an extra share to its marginal costs (price impact). The second part is most conveniently written as $((\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})(\mathbf{v_i}\mathbf{\Lambda}^{-1})$. In this form, it can be seen as a vector of shocks ($\mathbf{\Lambda}^{-1}$, resulting from the analysis above) multiplied with a matrix that

tells the investor how to allocate a shock. Not surprisingly, this allocation matrix looks very similar to what we have seen before, only this time multiplied with the unity vector to account for the fact that we have a vector of shocks rather than just one funding shock. The final part is the multiplication factor $\frac{1}{2W_i(1+Z_i)}$, which follows from the fact that for wealthy investors, less is to be gained in relative terms because transaction costs quickly outweigh informational advantages.

As before, we can obtain order flow by:

$$\mathbf{o_i} = W_i(1+Z_i)\mathbf{x_i} - W_i\mathbf{x_i^*} \tag{10}$$

$$= W_i Z_i \mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1} + \mathbf{\Lambda}^{-1}(\mathbf{I} - (\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})\mathbf{v_i}. \tag{11}$$

The private information induced component of the order flow can be interpreted as follows. First, the matrix $\mathbf{\Lambda}^{-1}$ dictates that the amount of trading on private information for a given security is inversely related to the price impact of trading volume, which is intuitive. Second, the matrix $(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})$ results from the budget constraint and reflects the proportions in which an information shock in one security is funded by each of the others. The rows of this matrix add up to one. Third, the setting is constructed such that each individual investor trades on information shocks in such a way that the transaction costs on a marginal dollar of trading are exactly equal to (and therefore offset by) the alpha gain. Thus, informed trading volume is independent of wealth.[6]

3.3. Aggregating to market level and extracting consensus information

Aggregating order flow across all investors gives:

$$\mathbf{o_m} = \sum_i \mathbf{o_i}$$

$$= \mathbf{\Lambda}^{-1}(\mathbf{I} - (\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})M\bar{\mathbf{v}} + \sum_i W_i Z_i \mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}, \tag{12}$$

[6]This assumption might be unrealistic as some of the small investors would have to go short heavily in some of their securities to fund their uninformed trading. An extra set of restrictions on non-negative holdings may resolve this issue, but leads to less tractable results that are harder to interpret.

where $\bar{\mathbf{v}}$ is the average (equally-weighted) information shock. In (12), $\mathbf{\Lambda}^{-1}M\bar{\mathbf{v}}$ refers to the aggregate speculative trading volume, $-\mathbf{\Lambda}^{-1}((\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})M\bar{\mathbf{v}}$ refers to the aggregate funding demand for the speculative trades and $\sum_i W_i Z_i \mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}$ refers to the aggregate liquidity demand. $M\bar{\mathbf{v}}$ can be thought of as the aggregate amount of private information (incidence rate times size) in the market.

In our attempts to obtain a measure of informed trading, we can try to invert (12) to end up with an analytical expression for $M\bar{\mathbf{v}}$. However, because we allow for an information shock for each security, the matrix $\mathbf{\Lambda}^{-1}(\mathbf{I}-(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})$ is not full rank and hence cannot be inverted. The reason for this can be seen in a two security example. Observing positive order imbalance for security 1 and negative order imbalance for security 2, could imply either (i) a positive information shock for security 1, which is associated with selling of security 2 to fund the speculative trade in security 1, or (ii) a negative information shock for security 2, leading to buying in security 1 with the funds received from selling security 2. These two are empirically indistinguishable. To solve our under-identification problem, we assume that one of our securities never suffers from informed trading. This can be a treasury bond or an information-insensitive security. In our implementation in Section 5, we use the SPDR S&P500 ETF. We refer to this security as the "benchmark security. "

When working out $M\bar{\mathbf{v}}$, we obtain a remarkably simple expression:

Proposition 1. *The order-flow implied aggregate private information shock for security* $j \in \{2, .., N\}$ *is given by:*

$$M\bar{v}_j = \lambda_j o_j - \lambda_1 o_1, \tag{13}$$

where security 1 is the benchmark security.

Proof. See appendix. □

Our model thus not only allows us to decompose a security's aggregate order flow into informed trading on the one hand and liquidity-motivated and funding-induced trading on the other hand, but also yields a very simple and intuitive expression for a security's

aggregate private information shock: its $\lambda \times OIB$, in excess of the same term for a benchmark security that is insulated from informed trading. In the remainder of the paper, we set out to estimate and validate these measures of informed trading and of the aggregate private information shocks for a large sample of U.S. stocks over a prolonged time period.

4. Data and variable definitions

For our empirical analysis of the model introduced in Sections 2 and 3, we use a sample of S&P 1500 stocks over 2001-2002. Our motivation for using S&P 1500 stocks is that most institutional investors focus on stocks with a relatively large market capitalization, so that this sample represents a reasonable set of stocks that informed traders might consider. The choice for S&P 1500 stocks also aims to strike a balance between ensuring a sample of sufficient breadth, while at the same time excluding small and thinly traded stocks for which the estimation of order imbalance and price impact parameters based on intraday data is problematic. 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 2010. We refer to Appendix C for a detailed description of the sample selection and composition.

All of our analyses are done at the daily frequency, where the key parameters (order imbalance and price impact) are estimated each day for each stock based on intraday data. We obtain intraday price and transaction data for individual stocks from the NYSE Trade and Quote (TAQ) database. To preclude survivorship bias, we obtain data for each stock over the entire period for which we have data over 2001-2010, and not only for the period during which they were an S&P 1500 constituent. We refer to Appendix D for a detailed description of the data screens and filters we apply to the TAQ data, all of which are taken from prior studies dealing with these data.

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 midpoint, we classify it as a buy (sell). If a trade occurs exactly at the quote mid-point, we sign it using the previous transaction 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 as the previous trade (a zero tick), then the trade is a zero-uptick if the previous price change was positive. If the previous price change was also equal to zero, we discard the trade. We do not use a delay between a trade and its associated quote because of the decline in reporting errors (see Madhavan, Richardson, and Roomans, 2002; 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 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 USD.

We estimate the daily price impact parameter for each stock using the approach of Glosten and Harris (1988), based on daily regressions of the price change of a trade relative to the previous trade on the current quantity traded and the change in the sign of the trade. The coefficient on the quantity traded 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. We scale the estimate of this coefficient by the squared closing price (quote midpoint) at the end of the same trading day to make sure that, in line with the model, price impact is measured as the percentage price change per unit of dollar trading volume.

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 the rebalancing of their portfolio following liquidity and private information shocks based on the expected price impact of trading different securities. In other words, estimating price impact parameters over the same day as we measure the order imbalances (that within the model arise as a result of the portfolio rebalancing by individual investors) would introduce look-ahead bias into our analyses. 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 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.

Our returns-based empirical analyses are based on midquote returns computed from the daily midpoint of the last quote on each day, adjusted for corporate actions using CRSP data, and cross-sectionally winsorized each day at the 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 midpoint of the quotes, averaging the spread across all trades for the stock on that day ($PQSPR$). 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 estimating these variables, we drop stocks with fewer than six months of data. 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 2,130 stocks (listed at NYSE, Nasdaq, or Amex) that were an S&P 1500 constituent at some point during our sample period of 2001-2010 and that survive these data screens.

We use the SPDR S&P500 ETF (ticker "SPY") as a benchmark security that is insulated from informed trading, which is needed to tackle underidentification of the model. Our motivations for choosing the SPDR as the benchmark security 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. We obtain consolidated trades and quotes for the SPDR from the Thomson Reuters Tick History (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.

5. Empirical results

The main purpose of our empirical analyses is to examine whether the measures of informed trading and private information stemming from the model developed in Sections 2 and 3 can be applied to real-life data and yield results that are consistent with our

theoretical interpretation of these measures.

For each stock on each day, we estimate the (signed) dollar volume of informed trading using the decomposition of the stock's aggregated order flow on that day into informed trading, liquidity trading, and funding trading, as expressed in equation (12). This expression is worked out in more detail in equation (A.21) in Appendix A. Solving for an individual stock's informed trading volume is based on our estimates of the order imbalance (OIB) and price impact parameter (λ) of the stock of interest, of all other S&P 1500 constituents in our sample on that day, and of the SPDR (our benchmark security for which we assume informed trading volume to be equal to zero) on that day. For ease of interpretation, we scale the absolute informed trading volume by total trading volume for that stock on that day. The resulting measure, which we label $XPIN$, can be interpreted as the propensity or probability of informed trading.

We also estimate the aggregate private information shock (or $M\bar{v}$) for each stock on each day based on equation (13). This measure of private information is based on just the estimates of the order imbalance and price impact parameter of the stock of interest and of the SPDR.

Table 1 presents summary statistics of the daily returns, OIB, λ, $PQSPR$, $XPIN$, and $M\bar{v}$ across all stocks in our sample over 2001-2010. 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 2,130 S&P 1500 constituents in the sample, for which we have daily observations for 1,829 days on average.

The mean and median mid-quote returns are equal to, respectively, five and six basis points per day. The median OIB is slightly positive ($0.18m.) over our sample, but, not surprisingly, exhibits substantial cross-sectional variation, with a standard deviation of $3.39m. The median λ (scaled by 10^6) equals 0.29%, which means that the median of the average price impact across all stocks in the sample is 29 basis points for a trade of $1m. The median $PQSPR$ is 20 basis points. The mean order imbalance and price impact estimate of the SPDR benchmark security are equal to, respectively, $17.29m. and 0.09 basis points per $1m trade (not tabulated), which indicates 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 1000th of the cross-sectional mean of the average price impact of the S&P 1500 stocks of 0.95%.

The mean and median $XPIN$ are equal to 0.15 and 0.16, respectively, which indicates that our approach identifies roughly 15% of the trading volume in individual stocks on a given day as informed. This number is comparable in magnitude to the mean and median PIN estimate of around 19% reported by Easley, Hvidkjaer, and O'Hara (2002).

The mean and median $M\bar{v}$ are equal to 0.09 and 0.04, respectively, which suggests that the aggregate private information shock was slightly positive in our sample. The magnitude of $M\bar{v}$ is difficult to interpret, since it requires an assumption about the number of investors (M). However, the sign of $M\bar{v}$ does indicate whether the aggregate private information shock was positive or negative for a given stock on a given day. Furthermore, the magnitude of $M\bar{v}$ can be compared across stocks in the sense that a greater $M\bar{v}$ indicates a greater aggregate private information shock. The cross-sectional standard deviation of the average $M\bar{v}$ of individual stocks is substantial, at 0.18.

To get a sense of the time-series variation in private information in our sample, we plot the average $M\bar{v}$ of the top and bottom decile portfolios of stocks sorted on $M\bar{v}$ each day in Figure 1. Consistent with the summary statistics in Table 1, the aggregate private information shock tends to be somewhat larger in magnitude for stocks with positive private information shocks than for stocks with negative private information. The degree of private information is relatively high for both decile portfolios in the first few years over our sample period, then decreases slowly over time in 2003-2007 (both for positive and negative shocks), after which it shows a peak again in the period surround the start of the financial crisis in 2008-2009, to return to pre-crisis levels by 2010.

Figure 2 provides a first indication of the relation between $M\bar{v}$ and contemporaneous stock returns by plotting the time-series of the returns of the top and bottom decile portfolios of stocks sorted on $M\bar{v}$ each day (from Figure 1). The patterns in Figure 2 are a near mirror image of those in Figure 2, which suggests that the contemporaneous returns of stocks with positive (negative) private information tend to be positive (negative) and that the strength of this relation is relatively stable over time.

In our empirical tests, we focus on our measure of the aggregate private information shock ($M\bar{v}$) rather than on our measure of the probability of informed trading ($XPIN$), for two reasons. First, our private information shock measure is signed and thus contains more information. Second, the predictions about the relation with the cross-section of returns are more clear-cut for the private informed measure than for the informed trading measure. For example, we would expect $M\bar{v}$ to be linearly related to contemporaneous stock returns, but for $XPIN$ it is less clear what to expect, because $XPIN$ is unsigned but also because $XPIN$ depends on the amount of liquidity-motivated trading and not only on the underlying information signal.

Furthermore, since all of our empirical tests are cross-sectional in nature, we can use a further simplified version of our private information measure: the product of a stock's estimated order imbalance and price impact ($\lambda \times OIB$). Because the correction for the benchmark's product of order imbalance and price impact in equation (13) is the same for all stocks on a given day, this simplification does not affect the results.

Table 2 shows the pooled contemporaneous correlations between $M\bar{v}$, the absolute value of $M\bar{v}$, $PQSPR$, λ, the further simplified private information measure ($\lambda \times OIB$), OIB, and $Return$. As expected, a stock's quoted spread is positively correlated to the absolute magnitude of private information in that stock as well to the stock's price impact. The order imbalance is negatively correlated with both $PQSPR$ and λ. $M\bar{v}$ is highly correlated with its simplified version $\lambda \times OIB$ (at 0.645), but not perfectly, which stems from time-series variation in the product of order imbalance and price impact of the benchmark security that will not influence our cross-sectional tests. We note that the correlations of both λ and OIB with $\lambda \times OIB$ are relatively small (at 0.013 and 0.159, respectively), 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. The correlations with returns provide some further initial evidence that our measures pick up meaningful variation in private information, since both $M\bar{v}$ and $\lambda \times OIB$ are positively and significantly related to contemporaneous stock returns. At around 0.10, these correlations are not overwhelming, but daily returns for individual stocks are noisy and we note that both correlations are more than double the

magnitude of the correlation between OIB by itself and contemporaneous returns.

In Table 3, we substantiate the initial evidence on the positive association between our private information measure $\lambda \times OIB$ and contemporaneous returns by running daily Fama-MacBeth (1973) regressions of the midquote returns on individual stocks on one-day lagged returns, OIB, λ, and $\lambda \times OIB$. OIB is included contemporaneously, since our approach aims to extract informed trading from the realized order imbalance on a given day. We note, however, that λ is not the contemporaneous price impact parameter for a stock on that day, but rather the expected price impact based on the moving average price impact estimates over the past 20 days (excluding the current day), since the model assumes that order flow on a given day is affected by the expected price impact of trading.

Consistent with prior studies, we find that daily stock returns exhibit a significantly negative autocorrelation. The coefficient on lagged returns is equal to -0.07 in the first model in Table 3, with a Fama-MacBeth t-stat of 14.7 (based on the Newey and West, 1987, correction for autocorrelation in the estimated coefficients). Not surprisingly, daily stock returns are significantly higher on days with more positive OIB. However, the interpretation of this finding is ambiguous, as both liquidity-motivated and informed trading are associated with price impact. The coefficient on λ is also positive and significant in most regression models in Table 1. This positive effect of λ on contemporaneous returns was not clear ex ante, but may be driven by the fact that the order imbalance is positive on average in our sample.

More importantly, we find a positive and highly significant effect of our simplified private information measure $\lambda \times OIB$ on contemporaneous returns. This result suggests that returns are 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 economic magnitude of this effect is considerable. A one standard deviation increase in $\lambda \times OIB$ is associated with a 0.12 standard deviation increase in contemporaneous stock returns, which is substantial in light of the noise inherent in daily stock returns. We note that the effect of our private information measure $\lambda \times OIB$ 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 well-known parts.

In the final two regression models of Table 3, 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 fourth model in Table 3, we include the product of OIB and the inverse of a stock's market capitalization. In the fifth model, we include the product of OIB and $PQSPR$. Although the coefficients of both $\lambda \times 1/Mktcap$ and $\lambda \times PQSPR$ are positive and significant, the effect of $\lambda \times OIB$ remains intact.

We next turn to potentially more stringent tests of our conjecture that $\lambda \times OIB$ measures private information. For this conjecture to be validated, we should observe significantly weaker return reversals following stock-days with large positive or negative values of $\lambda \times OIB$, since informed trading should be associated with permanent rather than transitory price impact. We test this hypothesis in two ways.

Table 4 reports the results of daily Fama-MacBeth regressions of the midquote returns on 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 have the opposite sign as the coefficient on lagged returns. We note that we take the absolute value of our private information measure $\lambda \times OIB$ for these tests, since return reversals should be weaker following large positive or negative information shocks. However, because λ is non-negative by construction, we only need to take the absolute value of OIB.

Consistent with Table 3, the first-order autoregressive coefficient is significantly negative, at -0.09 in the first model of Table 4. 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 extreme private information shock on the previous day.[7]

The coefficient on the interaction term of lagged returns and lagged $\lambda \times |OIB|$ is

[7]This effect may be driven by our finding in Figure 1 that over sample period positive information shocks tend to be somewhat greater than negative shocks. However, we note that the lagged effect of $\lambda \times |OIB|$ is much less significant in both statistical and economic terms compared to the contemporaneous effect of $\lambda \times OIB$ reported in Table 3, which is what we would expect.

significantly positive at 2.17, with a Fama-MacBeth Newey-West t-stat of 10.35. This finding indicates that, indeed, stock returns tend to revert significantly less following stock-days with high absolute values of our private information measure. We interpret this evidence as consistent with the view that $\lambda \times OIB$ does proxy for aggregate private information shocks. The third model of Table 4 shows that this result survives breaking up $\lambda \times |OIB|$ into its two separate variables and including all the relevant interactions.

To assess the economic significance of the reduced strength of return reversals following stock-days with high absolute values of $\lambda \times OIB$, we also analyze the returns on reversal strategies separately for stock-day observations with low and high private information. To that end, we first sort stocks into quintile portfolios on day $d-1$ based on their $\lambda \times OIB$. Quintile portfolios 1 and 5 thus contain stocks with, respectively, large negative and large positive 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 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. The difference between the abnormal returns on the reversal strategies within the low and high private information quintiles can be interpreted as a measure for how large the reduction in the strength of return reversals is following high $\lambda \times OIB$ stock-days.

The results of this second, 5×5 double-sorts approach to analyzing the strength of return reversals following low and high private information stock-days are in Panel A of Table 5. The first four columns of the panel report the estimates of time-series regressions of the daily returns on the reversal strategy for private information quintile 3 (which contains stocks whose aggregate private information estimate is close to zero) on various commonly used 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). We obtain daily returns on these factors from

the website of Ken French. All four models indicate economically large and statistically highly significant abnormal returns (alphas) of 46-47 basis points per day, which indicate strong daily return reversals for stocks with low private information estimates.[8]

The final column of Panel A shows the five-factor alpha of the difference between the reversal strategy for low private information stocks (private information quintile 3, as in columns 1-4) and the reversal strategy for high private information stocks (private information quintiles 1 and 5 combined). This alpha is significantly positive at 0.12 (Newey-West t-stat 4.81), which implies that the strength of return reversals is 12 basis points per day less following stock-days with high private information when compared to stock days following low private information, an effect that is significant from an economic perspective.[9]

In Panel B of Table 5, we reverse the 5×5 double sorting procedure by first sorting stocks into quintile portfolios based on their return on day $d-1$ and then sorting winner and loser stocks into five subportfolios based on their private information on day $d-1$. We then create an alternative reversal strategy that is long loser stocks with very positive values of $\lambda \times OIB$ and short winner stocks with very negative values of $\lambda \times OIB$. The idea is that stock-days with very positive private information but very negative returns or with very negative private but very positive returns are likely characterized by a large amount of liquidity-motivated trading in the opposite direction of the private information signal, and should thus exhibit strong reversals on the next day. The first four columns of Panel B show that the one-, three-, four-, and five-factor alphas of this strategy are economically and statistically large, at 35-37 basis points per day, with t-stats close to

[8]These abnormal return estimates on reversal strategies are somewhat higher than the mean reversal returns reported by Nagel (2012) of 18 basis points per day based on midquote returns and 30 basis points per day based on trade returns. This difference in magnitudes can likely be explained by differences in the sample, by the fact that Nagel's reversal strategy returns are based on all stocks rather than only the extreme winner and loser stocks, and by the fact that the first four models in Panel A of Table 5 use only stocks for which we estimate the amount of private information to be low. We note that neither one of these reversal strategy return estimates is realistic in the sense that they do not take into account transaction costs and short-sales constraints. We also note that we obtain qualitatively similar results when using trade returns instead of midquote returns.

[9]There are still significant return reversals following stock-days with high private information, but we note that our model does not rule out non-trivial liquidity-motivated trading on those stock-days, which could explain those return reversals.

10. The final column of Panel B compares the return on this strategy to the return on a reversal strategy that is long loser stocks with very negative values of $\lambda \times OIB$ and short winner stocks with very positive values of $\lambda \times OIB$, since the reversals should be weaker on these categories of stocks if $\lambda \times OIB$ is a meaningful proxy for private information. The significant difference in the abnormal returns on these two strategies of 12 basis points per day indicates that return reversals are considerably weaker when the returns on loser and winner stocks are more likely to be driven by private information.

Overall, our tests show that, consistent with our private information measure picking up meaningful cross-sectional variation in aggregate information shocks, stocks with a more positive value of $\lambda \times OIB$ tend to have significantly more positive contemporaneous returns, and stocks with very negative or very positive private information estimates subsequently exhibit significantly weaker return reversals. Both of these results support the theoretical interpretation of our new private information measure.

6. Conclusion

This paper proposes new measures of both the amount of informed trading in individual securities and the direction and magnitude of the aggregate private information shock for these securities. Both measures are derived from a portfolio optimization model for individual investors who are exposed to information and liquidity shocks. Our identification of informed trading is cross-sectional in the sense that it is based on the cross-section of price impact parameters and order imbalances for a given day (or intraday period).

We validate our private information measure by estimating it for all S&P 1500 stocks each day over 2001-2010. In particular, we show that it is strongly related to contemporaneous returns, and that return reversals are significantly weaker following stock-days with high private information estimates. Both pieces of evidence are consistent with the conjecture that our private information measure is indeed associated with the aggregate private information shock of individual securities.

An appealing feature of our private information measure is that it is intuitive and easy to estimate, even at high frequencies. In cross-sectional applications, it simplifies to a security's order imbalance multiplied by its price impact parameter ($\lambda \times OIB$). Fur-

thermore, in contrast to other measures that proxy for private information, our measure also conveys the direction of the private information signal. We hope that our measure will be useful in a host of applications in market microstructure, asset pricing, and corporate finance. In future work, we plan to investigate the asset pricing applications of our private information measure.

Appendix A. Proofs

Proof of Lemma 1

We solve the problem by a standard Lagrangian multiplier technique . We define

$$L(\mathbf{x_i},\mu) \quad = \quad \mathbf{x_i}'\iota\zeta - \frac{1}{1+Z_i}(W_i(\mathbf{x_i}(1+Z_i)-\mathbf{x_i^*})'\mathbf{\Lambda}(\mathbf{x_i}(1+Z_i)-\mathbf{x_i^*}))-\mu(\iota'\mathbf{x_i}-1). \quad \text{(A.1)}$$

The necessary FOCs for optimality are given by

$$\frac{\partial L(\mathbf{x_i},\mu)}{\partial \mathbf{x_i}} = 0, \qquad\qquad \frac{\partial L(\mathbf{x_i},\mu)}{\partial \mu} = 0. \qquad \text{(A.2)}$$

As $L(\mathbf{x_i},\mu)$ 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

$$\begin{bmatrix} -\mathbf{a_i} \\ 1 \end{bmatrix} = \begin{bmatrix} -\mathbf{Q_i} & \iota \\ \iota' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x_i} \\ \mu \end{bmatrix}, \qquad \text{(A.3)}$$

where

$$\mathbf{Q_i} = 2W_i(1+Z_i)\mathbf{\Lambda} \qquad \text{(A.4)}$$

$$\mathbf{a_i} = \iota\zeta + 2W_i\mathbf{\Lambda}\mathbf{x_i^*}. \qquad \text{(A.5)}$$

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

$$\mu = -(\iota'\mathbf{Q_i}^{-1}\iota)^{-1}\iota'\mathbf{Q_i}^{-1})\mathbf{a_i} + (\iota'\mathbf{Q_i}^{-1}\iota)^{-1} \qquad \text{(A.6)}$$

$$\mathbf{x_i} = \mathbf{Q_i}^{-1}(\mathbf{I} - \iota(\iota'\mathbf{Q_i}^{-1}\iota)^{-1}\iota'\mathbf{Q_i}^{-1})\mathbf{a_i} + \mathbf{Q_i}^{-1}\iota(\iota\mathbf{Q_i}^{-1}\iota')^{-1}. \qquad \text{(A.7)}$$

$$= \mathbf{Q_i}^{-1}\mathbf{a} + \mathbf{Q_i}^{-1}\iota\mu. \qquad \text{(A.8)}$$

If we define $f = (\iota' \mathbf{Q_i}^{-1} \iota)^{-1}$, we can work out μ:

$$\mu = -f\iota' \mathbf{Q_i}^{-1}(\iota\zeta + 2W_i \mathbf{\Lambda x_i^*}) + f \tag{A.9}$$

$$= -f(\iota' \mathbf{Q_i}^{-1}\iota)\zeta - f\iota' \mathbf{Q_i}^{-1} 2W_i \mathbf{\Lambda x_i^*} + f. \tag{A.10}$$

Substituting back f gives

$$\mu = -\zeta - f\iota' \mathbf{Q_i}^{-1} 2W_i \mathbf{\Lambda x_i^*} + f. \tag{A.11}$$

Substituting $\mathbf{Q_i}$ back gives

$$\mu = -\zeta - f\iota' \frac{1}{1+Z_i} \mathbf{x_i^*} + f. \tag{A.12}$$

Realizing that $\iota' \mathbf{x_i^*} = 1$ and multiplying f with $\frac{1+Z_i}{1+Z_i}$ gives

$$\mu = -\zeta + f\frac{Z_i}{1+Z_i}. \tag{A.13}$$

Now working out (A.8) gives

$$\mathbf{x_i} = \mathbf{Q_i}^{-1}\iota\zeta + \mathbf{Q_i}^{-1} 2W_i \mathbf{\Lambda x_i^*} - \mathbf{Q_i}^{-1}\iota\zeta + \mathbf{Q_i}^{-1}\iota f \frac{Z_i}{1+Z_i} \tag{A.14}$$

$$= \frac{1}{1+Z_i}\mathbf{x_i^*} + \frac{Z_i}{1+Z_i}\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}. \tag{A.15}$$

Proof of Lemma 2

With information shocks, (1) changes to

$$\begin{bmatrix} -\mathbf{a_i^v} \\ 1 \end{bmatrix} = \begin{bmatrix} -\mathbf{Q_i} & \iota \\ \iota' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x_i} \\ \mu \end{bmatrix}, \tag{A.16}$$

where

$$\mathbf{Q_i} = 2W_i(1+Z_i)\mathbf{\Lambda} \tag{A.17}$$

$$\mathbf{a_i^v} = \iota\zeta + 2W_i\mathbf{\Lambda}\mathbf{x_i^*} + \mathbf{v_i}. \tag{A.18}$$

Working through, we get the solution

$$\mathbf{x_i} = \frac{1}{1+Z_i}\mathbf{x_i^*} + \frac{Z_i}{1+Z_i}\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1} + \frac{1}{2W_i(1+Z_i)}\mathbf{\Lambda}^{-1}(\mathbf{I} - (i'\mathbf{\Lambda}^{-1}i)^{-1}ii'\mathbf{\Lambda}^{-1})\mathbf{v_i}. \tag{A.19}$$

Proof of Proposition 1

Equation (12) writes like (using H, the harmonic average lambda):

$$\mathbf{o_m} = \mathbf{\Lambda}^{-1} \times \left[1 - \begin{pmatrix} \frac{H}{N\lambda_1} & \frac{H}{N\lambda_2} & \cdots & \frac{H}{N\lambda_N} \\ \frac{H}{N\lambda_1} & \frac{H}{N\lambda_2} & & \frac{H}{N\lambda_N} \\ \vdots & & \ddots & \vdots \\ \frac{H}{N\lambda_1} & \frac{H}{N\lambda_2} & \cdots & \frac{H}{N\lambda_N} \end{pmatrix} \right] \times M\bar{\mathbf{v}} + \begin{pmatrix} \frac{H\sum_j o_j}{N\lambda_1} \\ \frac{H\sum_j o_j}{N\lambda_2} \\ \vdots \\ \frac{H\sum_j o_j}{N\lambda_N} \end{pmatrix}, \tag{A.20}$$

where $\bar{v}_1 = 0$. We hence have:

$$o_j = \frac{H\sum_j o_j}{N\lambda_j} + \lambda_j^{-1}Mv_j\left(1 - \frac{H}{N\lambda_j}\right) - \lambda_j^{-1} \times HM\sum_{-j}\frac{v_k}{N\lambda_k}, \tag{A.21}$$

or:

$$\frac{N\lambda_j}{H}o_j - \sum_k o_k = \frac{NMv_j}{H} - M\sum_k \frac{v_k}{\lambda_k} \tag{A.22}$$

or (by subtracting this equation for $j = 1$ from the equation for any other j):

$$Mv_j = \lambda_j o_j - \lambda_1 o_1. \tag{A.23}$$

Appendix B. Overview of notation used

Parameters		
Symbol	*Support*	*Description*
ζ	$\mathbb{R}$	Market price of risk
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
W_i	$\mathbb{R}^+$	Starting wealth of investor i
Z_i	$\mathbb{R}$	Liquidity shock of investor i
$\mathbf{x_i^*}$	$\mathbb{R}^N$	Starting portfolio of investor i
Indices		
i	$\{1,..,M\}$	Investors
j	$\{1,..,N\}$	Securities
Decision variables		
$\mathbf{x_i}$	$\mathbb{R}$	Portfolio allocation of investor i
$\delta_{i,j}$	$\{-1,1\}$	Trading direction of investor i in security j
$\mathbf{\Delta_i}$	$\mathbb{Z}^N\times\mathbb{Z}^N$	Diagonal matrix containing all $\delta_{i,j}$s of investor i
$\mathbf{o_i}$	$\mathbb{R}^N$	Order flow of investor i

Appendix C. Sample selection and composition

This appendix describes the selection and composition of our sample of S&P 1500 stocks. Our starting point is a list of all 2,553 stocks that were a constituent of the S&P 1500 index at some point in the period from January 2001 till December 2010 (including tickers, CUSIPs, and begin and end dates of the S&P 1500 index membership) – downloaded on February 3, 2011 from Compustat Monthly Updates North America Index Constituents. There are 2,392 unique tickers in this list.[10] As TAQ is organized by ticker (or symbol in TAQ terminology), we use the TAQNAMES file downloaded on January 1, 2010 (and for later years the monthly TAQ Master files for December 2009 and December 2010 downloaded on 27 July 2011, as TAQNAMES is no longer available) to check whether the Compustat tickers are available in TAQ. Of the 2,392 unique tickers, 346 cannot be found in TAQ. For the stocks with these tickers, we check whether an adjusted ticker that refers to the same stock is available in TAQ (based on a comparison of the 8-digit CUSIP and/or stock name on Compustat and TAQ). We make adjustments to 331 of the tickers. We note that most of these adjustments are trivial, such as removing "." or ".1" at the end of the ticker. We discard 15 stocks for which we could not find a corresponding ticker in TAQ. As we want to analyze only stocks listed on NYSE, AMEX, or Nasdaq, we obtain a list of all Compustat stocks and their stock exchange (data item EXCHG – which is the most recent exchange the stock was listed on) – downloaded on May 26, 2011 from Compustat. We also need the exchange of each stock because we follow prior studies and only download quotes for each stock from their own exchange. If the exchange in this list

[10] In most cases, multiple identical tickers occur on the list when the same stock (same name and same 8-digit CUSIPs) is listed as an S&P 1500 index constituent multiple times. In several cases, these different entries refer to distinct periods of S&P 1500 membership (such as Ace Ltd., which has entries for the period from January 30, 2002 till July 17, 2008 and for the period from July 15, 2010 till the end of our sample period). However, in a substantial number of cases, the different entries refer to consecutive periods of S&P 1500 membership for the same stock, with at most one trading day—quite often this day is Friday, August 1, 2003—in between the periods (such as U.S. Steel, which has entries for the period from February 1, 2001 till July 31, 2003, from August 4, 2003 till August 28, 2005, and from August 29, 2005 till the end of our sample period). We treat these consecutive periods with at most one trading day in between as one continuous index membership period. When the different S&P 1500 membership periods for a particular stock are non-consecutive, we download the entire data history available in TAQ for those stocks, though we later retain only the longest uninterrupted period for stocks for which there is a gap in the data of more than two months.

does not equal NYSE, AMEX, or Nasdaq, we manually check (primarily using internet searches) whether the stock was listed on one of these exchanges in an earlier period. Most stocks on our S&P 1500 constituents list for which Compustat indicates a different exchange than these three are stocks that went into bankruptcy or went private but used to be listed. For stocks that change from one of these three exchanges to another one of these three exchanges during our sample period, we only use the data for the most recent exchange the stock was listed on. After this procedure, there are 2,342 unique adjusted tickers, for which we download and process intraday TAQ data over the period 2001-2010 to construct daily measures of order imbalance and price impact. As the same ticker can be used on TAQ by multiple stocks in different periods, it is important to check whether the downloaded TAQ data for each ticker actually corresponds to the same stock in our list of S&P 1500 constituents. To that end, we look up each ticker in our list of S&P 1500 stocks in the TAQNAMES and/or TAQ Master files and verify that it is the same stock based the stock name, the 8-digt CUSIP, and the begin and end dates of the presence of the stock on TAQ. This verification has to be carried out manually, because TAQNAMES often contains different rows for the same ticker and even the same stock. If a stock's ticker is not in our TAQNAMES file (which covers the period till the end of 2008), we check whether it is in the TAQ Master files of December 2009 and/or December 2010. If that is the case, we use the start and end of those years as the begin and end dates on TAQ, realizing that TAQ data may not be available over those full years. If the period during which a stock appears on TAQ does not overlap with the period during which it is an S&P 1500 constituent, we discard the stock.[11] In line with the recommendation of WRDS, we use the 8-digit CUSIP to match the TAQ data with CRSP based on the historical CUSIP (data item "NCUSIP") in CRSP and obtain the CRSP "PERMNO"

[11] In a small number of cases, the TAQ CUSIP is different from the Compustat CUSIP (usually only the seventh digit, which identifies the exact issue – where the first six digits identify the issuer), but the stock name and period correspond and there are no other stocks with the same symbol in TAQNAMES. In these cases, we retain the TAQ CUSIP, as this is the historical CUSIP that corresponds to the data we downloaded from TAQ for that stock. In some cases, TAQNAMES shows multiple lines for the same ticker with the same name and the same 8-digit CUSIP. If the begin and end dates of those different lines are consecutive, we treat them as representing a single stock. If not, and if TAQ only covers the period listed on one of the lines, we use that period. If one of the lines lists a longer period on TAQ that encompasses the shorter period listed on the other line, we use the longer period.

identifier for each stock in our list. We manually check whether the names in CRSP match those of our list of stocks, and whether different names refer to the same stock using the PERMNO and/or internet searches. We discard one stock for which we cannot find a match on CRSP. The resulting dataset consists of 2,302 different stocks (with 2,282 unique adjusted tickers), of which 1,408 are NYSE listings, 12 are AMEX listings, and 882 are Nasdaq listings. We note that we discard some more stocks based on further data screens discussed in Section 4 and Appendix D.

Appendix D. Data screens and filters applied to the TAQ data

This appendix describes the data screens and filters we apply to our sample of S&P 1500 stocks. We follow Hasbrouck (2007) and set the price of the first trade on a day to missing to cope with issues surrounding overnight price changes and special features of the opening. We discard bid and ask quotes that are less than or equal to 0, bid and ask sizes that are less than or equal to 0, and quote conditions (mode) that are not in 4, 7, 9, 11, 13, 14, 15, 19, 20, 27, 28, following WRDS recommendations. We only retain quotes from the primary listing exchange of each stock, but we use trades from all trading venues, not just the primary listing exchange, following Hasbrouck (2005). We discard trades that are out of sequence (as indicated by a sale condition that is in O, Z, B, T, L, G, W, J, K, following WRDS recommendations), recorded before the market open or after the market close (following Chordia, Roll, and Subrahmanyam, 2001), with special settlement conditions (as indicated by a correction indicator that is not in 0,1,2), or with a price less than or equal to 0 or a trade size less than or equal to 0, again following WRDS recommendations. We also discard trades with (i) a quoted spread less than \$0 or greater than \$5, (ii) a ratio of effective spread to quoted spread greater than 4, or (iii) a ratio of proportional effective spread to proportional quoted spread greater than 4 (following Chordia, Roll, and Subrahmanyam, 2001).

References

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

Baker, M. and J. C. Stein (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets 7*, 271–299.

Bongaerts, D., F. De Jong, and J. Driessen (2011). Derivative Pricing with Liquidity Risk: Theory and Evidence from the Credit Default Swap Market. *Journal of Finance 66*, 203–240.

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

Chordia, T., R. Roll, and A. Subrahmanyam (2001). Market Liquidity and Trading Activity. *Journal of Finance 56*, 501–530.

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

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*, 255–67.

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

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 (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.

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.

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

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

Hasbrouck, J. (2007). *Empirical Market Microstructure*. Oxford University Press.

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

Jegadeesh, N. (1990). Evidence of Predictable Behavior of Security Returns. *Journal of Finance 45*, 881–98.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica 53*, 1315–35.

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

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

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.

Manove, M. (1989). The Harm from Insider Trading and Informed Speculation. *Quarterly Journal of Economics 104*, 823–45.

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

Newey, W. and K. West (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies 61*, 631–653.

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

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

Table 1 – Cross-sectional summary statistics of time-series averages
This table reports the cross-sectional (across the 2,130 S&P1500 stocks in the sample) mean, standard deviation, first quartile, median, and third quartile of the time-series average by stock of the daily return from corporate action adjusted end-of-day mid-quotes winsorized at the 0.1% level (*Return*), the daily average proportional quoted spread ($PQSPR$), the price impact defined as the percentage return in prices due to a trading volume of \$1m. (λ, each day cross-sectionally winsorized at the 95% level), the daily difference between the total dollar volume of trades initiated by buyers and sellers (order imbalance in \$m.) (OIB), the ratio of daily aggregate informed trading over daily trading volume ($XPIN$), and the daily aggregate private information ($M\bar{v}$) from Eq. (13). The first column indicates the number of stocks over which the summary statistics are computed. The second column indicates the number of days the average stock is in the sample. The sample includes all 2,130 stocks (listed at NYSE, Nasdaq, or Amex) that were an S&P 1500 constituent at some point during our sample period of 2001-2010. Data to compute all variables in the table are from TAQ. The factor to adjust daily closing mid-quote data for corporate actions is from CRSP.

	#Stocks	Days	mean	stddev	25%	median	75%
Return [%]	2,130	1,829	0.05	0.10	0.02	0.06	0.09
$PQSPR$ [%]	2,130	1,829	0.37	0.79	0.12	0.20	0.39
λ	2,130	1,829	0.95	1.76	0.10	0.29	0.99
OIB	2,130	1,829	1.31	3.39	-0.01	0.18	1.21
$XPIN$	2,130	1,829	0.16	0.07	0.12	0.15	0.18
$M\bar{v}$ [%]	2,130	1,829	0.09	0.18	-0.01	0.04	0.15

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

This table reports pooled Pearson correlation coefficients between seven daily stock-specific variables: Aggregate private information ($M\bar{v}$), absolute private information ($|M\bar{v}|$), proportional quoted spread ($PQSPR$), price impact (λ), dollar order imbalance (OIB), the product of dollar order imbalance and price impact ($\lambda \times OIB$), and returns ($Return$). We refer to Table 1 for a description of these variables. Data to compute the variables are from TAQ and CRSP. P-values are in parentheses.

	$M\bar{v}$	$\lvert M\bar{v}\rvert$	$PQSPR$	λ	$\lambda \times OIB$	OIB	$Return$
$M\bar{v}$	1.000						
$\lvert M\bar{v}\rvert$	0.262 (0.00)	1.000					
$PQSPR$	-0.027 (0.00)	0.052 (0.00)	1.000				
λ	-0.002 (0.00)	0.003 (0.00)	0.187 (0.00)	1.000			
$\lambda \times OIB$	0.645 (0.00)	0.144 (0.00)	-0.020 (0.00)	0.013 (0.00)	1.000		
OIB	0.253 (0.00)	0.093 (0.00)	-0.032 (0.00)	-0.003 (0.00)	0.159 (0.00)	1.000	
$Return$	0.105 (0.00)	0.054 (0.00)	-0.005 (0.00)	0.003 (0.00)	0.100 (0.00)	0.050 (0.00)	1.000

Table 3 – Daily Fama-MacBeth regressions of returns on contemporaneous private information

This table reports the time-series averages of the estimated slope coefficients from daily cross-sectional regressions to explain differences in mid-quote returns across stocks. The dependent variable is the end-of-day mid-quote price 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 parameter 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 for stock i on day $d-1$ ($PQSPR_{i,d-1}$), and various interaction terms. Fama-MacBeth t-statistics are in parentheses using Newey-West corrections. Data to compute the variables are from TAQ. Market capitalization data as well as the factor to adjust prices by corporate actions are from CRSP. Some coefficients have been scaled for ease of presentation.

	Dependent variable: $Return_{i,d}$				
$Return_{i,d-1}$	-0.07	-0.07	-0.07	-0.07	-0.07
	(-14.71)	(-14.95)	(-17.10)	(-18.35)	(-17.36)
$OIB_{i,d}$ $\times 10^4$	0.96	0.97	0.14	0.01	-0.28
	(9.34)	(9.30)	(4.03)	(0.40)	(-4.70)
$\lambda_{i,d}$ $\times 10^2$		0.99	0.98	0.20	1.25
		(5.44)	(6.40)	(1.34)	(6.72)
$\lambda_{i,d} \times OIB_{i,d}$			0.39	0.30	0.37
			(21.73)	(23.18)	(18.79)
$1/Mktcap_{i,y-}$				164.83	
				(8.34)	
$OIB_{i,d} \times 1/Mktcap_{i,y-}$				549.36	
				(12.59)	
$PQSPR_{i,d-1}$					-0.01
					(-0.84)
$OIB_{i,d} \times PQSPR_{i,d-1}$					0.07
					(7.76)
R^2	2.43	2.86	4.95	6.13	5.70
# regressions	2,441	2,441	2,441	2,192	2,441

Table 4 – 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 predictive, cross-sectional regressions to explain differences in mid-quote returns across stocks. The dependent variable is the end-of-day mid-quote price 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 absolute order imbalance of stock i on day $d-1$ ($|OIB_{i,d-1}|$), the price impact parameter of stock i 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 ($\lambda_{i,d-1}$), and various interaction terms. Fama-MacBeth t-statistics are in parentheses using Newey-West corrections. Data to compute the variables are from TAQ. The factor to adjust prices by corporate actions is from CRSP. Some coefficients have been scaled for ease of presentation.

Dependent variable: $Return_{i,d}$			
$Return_{i,d-1}$	-0.09	-0.10	-0.10
	(-11.01)	(-12.34)	(-11.56)
$\lambda_{i,d-1}\times\|OIB_{i,d-1}\|$		0.04	0.04
		(6.32)	(6.46)
$Return_{i,d-1}\times\lambda_{i,d-1}\times\|OIB_{i,d-1}\|$		2.17	2.40
		(10.35)	(11.22)
$\|OIB_{i,d-1}\|\ \times 10^4$			-0.01
			(-5.28)
$Return_{i,d-1}\times\|OIB_{i,d-1}\|$			0.00
			(9.91)
$\lambda_{i,d-1}\ \times 10^2$			0.21
			(1.15)
$Return_{i,d-1}\times\lambda_{i,d-1}$			-0.37
			(-5.51)
R^2	2.60	3.45	5.05
# regressions	2,441	2,440	2,440

Table 5 – The returns on reversal strategies conditional on private information
This table reports the results of time-series regressions of factor models to explain profits from two different investment strategies, based on a double-sorting approach. In Panel A, we sort all stocks in our sample into five portfolios based on $\lambda \times OIB$ on day $d-1$. We then sort all stocks in the median $\lambda \times OIB$ portfolio into five subportfolios based on their return on day $d-1$. The dependent variable in Panel A is the equally-weighted return on day d of going long the "losers" (i.e., the bottom quintile portfolio sorted by past returns) and short the "winners" (i.e., the top quintile portfolio) within the median $\lambda \times OIB$ portfolio. In Panel B, we sort all stocks in our sample into five portfolios based on their return on day $d-1$. We then sort all stocks in the "winner" and "loser" portfolio into five subportfolios based on $\lambda \times OIB$ on day $d-1$. The dependent variable in Panel B is the equally-weighted return on day d of going long the high $\lambda \times OIB$ stocks in the "loser" portfolio and short the low $\lambda \times OIB$ stocks in the "winner" portfolio. Independent variables in the regressions are: 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 last columns in both Panel A and Panel B report the results of investing in the above strategies and subtracting the profits following a reversal strategy in the "opposite" $\lambda \times OIB$ portfolio, called a "control" strategy. In Panel A, the "control" strategy is going long the "losers" and short the "winners" in the two extreme $\lambda \times OIB$ portfolios. In Panel B, the "control" strategy is going long the low $\lambda \times OIB$ stocks in the "loser" portfolio and short the high $\lambda \times OIB$ stocks in the "winner" portfolio. Newey-West t-statistics are in parentheses. Data to compute the variables are from TAQ. The factor to adjust prices by corporate actions is from CRSP. Daily factor portfolio returns are from the website of Ken French.

Panel A: Return reversal in median information portfolio

					REV - Control
Intercept	0.46	0.47	0.47	0.46	0.12
	(5.56)	(6.87)	(7.61)	(7.67)	(4.81)
Mkt - RF	0.08	0.08	0.11	0.08	-0.07
	(3.98)	(4.09)	(6.10)	(4.16)	(-2.29)
SMB		-0.12	-0.14	-0.11	-0.08
		(-2.42)	(-2.77)	(-2.49)	(-1.85)
HML		-0.01	0.00	0.03	0.09
		(-0.34)	(0.05)	(0.60)	(2.11)
Momentum			0.08	0.08	-0.02
			(3.12)	(2.76)	(-0.57)
Reversal				0.13	-0.06
				(3.98)	(-1.51)
# Obs.	2,453	2,453	2,453	2,453	2,453
R^2	1.02	1.50	2.03	3.32	2.12

Table 5 – continued

Panel B: Return reversal in extreme return portfolios

					REV - Control
Intercept	0.36	0.37	0.36	0.35	0.11
	(10.52)	(11.93)	(9.99)	(11.34)	(4.28)
Mkt - RF	0.15	0.16	0.20	0.15	0.02
	(4.85)	(4.99)	(6.25)	(5.66)	(0.76)
SMB		-0.21	-0.22	-0.19	-0.23
		(-4.24)	(-4.49)	(-4.06)	(-4.65)
HML		-0.06	-0.04	-0.00	0.05
		(-0.98)	(-0.71)	(-0.08)	(1.01)
Momentum			0.10	0.10	0.01
			(2.85)	(2.56)	(0.37)
Reversal				0.17	-0.05
				(3.33)	(-1.49)
# Obs.	2,453	2,453	2,453	2,453	2,453
R^2	3.23	4.42	5.09	6.96	2.00

Figure 1 – Time-series of the average $M\bar{v}$ of the top 10% and the bottom 10% of all stocks sorted by $M\bar{v}$.
This figure shows monthly time-variation in the equally-weighted, aggregate private information ($M\bar{v}$) of the 10% of all stocks with the highest and lowest private information on each given day. Aggregate private information is defined as in Eq. (13). Data to compute $M\bar{v}$ is from TAQ.

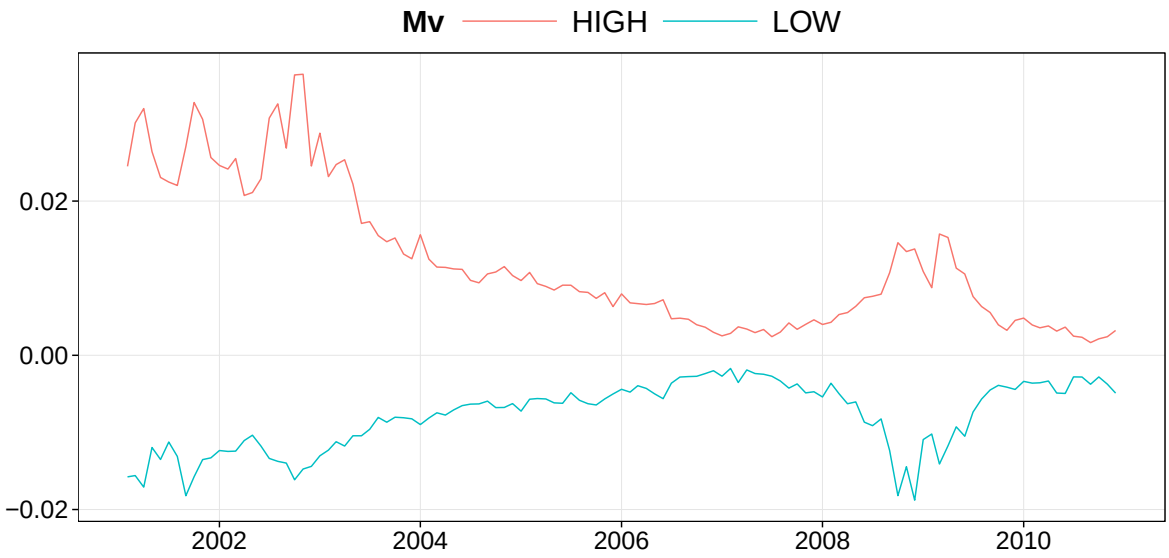

Figure 2 – Time-series of the average return of the top 10% and the bottom 10% of all stocks sorted by $M\bar{v}$.
This figure shows monthly time-variation of the end-of-day equally-weighted, mid-quote returns of the stocks in the top and bottom decile aggregate private information ($M\bar{v}$) portfolio. Aggregate private information is defined as in Eq. (13). Data to compute $M\bar{v}$ is from TAQ. The factor to adjust prices by corporate actions is from CRSP.