Identifying Information Asymmetry in Securities Markets

Kerry Back
Jones Graduate School of Business and School of Social Sciences,
Rice University

Kevin Crotty
Jones Graduate School of Business, Rice University

Tao Li
City University of Hong Kong

We propose and estimate a model of endogenous informed trading that is a hybrid of the PIN and Kyle models. When an informed trader trades optimally, both returns and order flows are needed to identify information asymmetry parameters. Empirical relationships between parameter estimates and price impacts and between parameter estimates and stochastic volatility are consistent with theory. We illustrate how the estimates can be used to detect information events in the time series and to characterize the information content of prices in the cross-section. We also compare the estimates to those from other models on various criteria. (*JEL* D82, D83, G14)

Information asymmetry is a fundamental concept in economics, but its estimation is challenging because private information is generally unobservable. Many proxies for information asymmetry exist including bid/ask spreads, price impacts, and estimates from structural models. In this paper, we study the identification of information asymmetry parameters in structural models. Structural modeling allows the econometrician to capture parameters

related to the underlying economic mechanisms such as the probability and magnitude of private information events or the intensity of liquidity trading. Demand for plausible measures of information asymmetry is high because private information plays a key role in so many economic settings. Evidence of this demand is the large literature in finance and accounting that utilizes the probability of informed trade (PIN) measure of Easley et al. (1996) to proxy for information asymmetry.[1]

Our first contribution is to propose and solve a model of informed trading in securities markets that shares many features of the PIN model of Easley et al. (1996) but in which informed trading is endogenous like in Kyle (1985). We call this a hybrid PIN-Kyle model. In the paper, we study a binary signal following Easley et al. (1996), but the model can accommodate more general signal distributions.

An important implication of the model is that order flows alone cannot identify information asymmetry. The intuition is quite simple. Consider, for example, a stock for which there is a large amount of private information and another for which there is only a small amount of private information. If it is anticipated that private information is more of a concern for the first stock than for the second, then the first stock will be less liquid, other things being equal. The lower liquidity will reduce the amount of informed trading, possibly offsetting the increase in informed trading due to greater private information. In equilibrium, the amount of informed trading may be the same in both stocks, despite the difference in information asymmetry. In general, the distribution of order flows need not reflect the degree of information asymmetry when liquidity providers react to information asymmetry and informed traders react to liquidity. Thus, we provide the first theoretical explanation of why methodologies that use order flows alone to estimate information asymmetry parameters, like PIN and Adjusted PIN (Duarte and Young, 2009), may not identify private information.[2]

---

[1] Some of those papers assess whether information risk is priced. See, for example, Easley and O'Hara (2004), Duarte and Young (2009), Mohanram and Rajgopal (2009), Easley, Hvidkjaer, and O'Hara (2002), Easley, Hvidkjaer, and O'Hara (2010), Akins, Ng, and Verdi (2012), Li et al. (2009), and Hwang et al. (2013). Many other papers use PIN (and other measures) to capture a firm's information environment in a variety of applications ranging from corporate finance (e.g., Chen, Goldstein, and Jiang, 2007, Ferreira and Laux, 2007) to accounting (e.g., Frankel and Li, 2004, Jayaraman, 2008).

[2] Several papers argue that PIN does not identify private information. Aktas et al. (2007) examine trading around merger announcements. They show that PIN decreases prior to announcements. In contrast, percentage spreads and the permanent price impact of trades, measured like in Hasbrouck (1991), rise before announcements, indicating the presence of information asymmetry. They describe the decline in PIN prior to announcements as a PIN anomaly. Akay et al. (2012) show that PIN is higher in the Treasury-bill market than it is in markets for individual stocks. Given that it is very doubtful that informed trading in Treasury bills is a frequent occurrence, this is additional evidence that PIN is not measuring information asymmetry. Benos and Jochec (2007) find that PIN is higher following earnings announcements, contrary to their assumption that information asymmetry should be higher before announcements. Duarte, Hu, and Young (2017) examine opportunistic insider trades. They estimate the parameters of the PIN model and then compute the conditional probability of an information event each day. They show that the conditional probability rises prior to opportunistic insider trades but stays elevated for a number of days following announcements. They argue that high turnover is misidentified as private information by the PIN model.

Our second contribution is to develop novel estimates characterizing the information environment in financial markets. We structurally estimate our theoretical model for a panel of stocks and provide several validation checks that the estimated parameters are plausibly related to information asymmetry. First, reduced-form estimates of price impact are increasing in our structural estimates of the probability and magnitude of information events, as implied by theory. Second, the model implies that the magnitude of price changes is proportional to Kyle's lambda, which depends on order flows and parameters of the model. Empirically, volatility over the latter part of a trading day is increasing in the conditional model-implied lambda, where the conditioning is based on cumulative order flows over the first part of the day and our estimated parameters. This phenomenon of stochastic volatility occurs in both the model and the data.[3]

To demonstrate potential applications of the estimates, we revisit two settings in which PIN estimates have been employed. One application of PIN has been to attempt to capture time-series variation in information asymmetry.[4] We show that conditional probabilities of information events calculated using order flows and our parameter estimates rise on average around earnings announcements and are higher both pre- and post-announcement for announcements with larger absolute earnings surprises. Private information is more likely to be present around such announcements. Conditional probabilities are also elevated during block accumulations by Schedule 13D filers, which existing information asymmetry measures fail to detect (Collin-Dufresne and Fos, 2015). These results indicate that the model does capture time-series variation in information asymmetry.

The second application illustrates how estimates of the information asymmetry parameters from our model can be used to augment studies concerned with cross-sectional differences in the information content of prices. To do so, we consider the hypothesis of Chen, Goldstein, and Jiang (2007) that corporate investment is more sensitive to market prices when there is more private information in prices. Our model allows us to measure the amount of private information alternatively by the frequency of private information events, by the magnitude of private information, and by the fraction of total price movement that is due to private information. We show that corporate investment is more sensitive to prices when any of these measures is higher. These measures of private information should prove useful in other settings in

---

[3] Banerjee and Green (2015) solve a rational expectations model with myopic mean-variance investors in which investors learn whether other investors are informed. They show that variation over time in the perceived likelihood of informed trading induces volatility clustering. While their model is quite different from ours, our model also exhibits volatility clustering. Volatility follows the same pattern as Kyle's lambda, which varies over time because of variation in the market's estimate of whether an information event occurred.

[4] For example, Brown, Hillegeist, and Lo (2004, 2009) examine changes in information asymmetry following voluntary conference calls and earnings surprises, respectively, and Duarte et al. (2008) study the effect of Regulation FD on PIN and the cost of capital.

which researchers are interested in capturing distinct facets of the information environment (e.g., the amount of liquidity trading or the magnitude of private information).

Related structural models of informed trading include the Adjusted PIN (APIN) model of Duarte and Young (2009), the Volume-Synchronized PIN (VPIN) model of Easley, López de Prado, and O'Hara (2012), and the modified Kyle model of Odders-White and Ready (2008). The APIN model allows for time variation in liquidity trading (with positively correlated buy and sell intensities), which provides a better fit to the empirical distribution of buys and sells. The VPIN model estimates buys and sells within a given time interval by assigning a fraction of total volume to buys and the remaining fraction to sells based on standardized price changes during the time interval.[5] Odders-White and Ready (2008; OWR) analyze a Kyle model in which the probability of an information event is less than 1, as it is in our model. However, they analyze a single-period model, whereas we study a dynamic model. Unlike our dynamic model in which prices equal conditional expectations, market makers in their model only match unconditional means of prices to unconditional means of asset values.[6]

Our estimate of the probability of an information event is not positively correlated in the cross-section with estimates from the other models. The divergence between the estimates is not surprising, because the models have different assumptions/implications regarding what data is required to identify the probability of an information event.[7] We also calculate a composite measure of information asymmetry in our model: the expected average lambda. This measure incorporates both the probability and the magnitude of information events, as well as the amount of liquidity trading. Unlike the probability of an information event, the expected average lambda from our model is positively correlated with similar measures from other models (PIN, APIN, VPIN, and the OWR lambda). Each of these measures should be increasing in the probability of an information event, so it is surprising that they are all positively correlated, given the lack of correlation of the 'probability of an information event' estimates. However, the measures are also decreasing in the amount of liquidity trading, and we present evidence in Section 4 that the measurement of liquidity

---

[5] Easley, López de Prado, and O'Hara (2011) claim that VPIN predicted the "flash crash" of May 6, 2010. This claim and some other claims regarding VPIN are challenged by Andersen and Bondarenko (2014b). See also Easley, López de Prado, and O'Hara (2014) and Andersen and Bondarenko (2014a).

[6] In a single-period model, because of the net order having a mixture distribution, the conditional expectation of the asset value given the net order is not a linear function of the net order. We solve our model by exploiting the local linearity of continuous time. Odders-White and Ready (2008) instead deviate from the usual Kyle model hypothesis that prices equal conditional expected values and instead find a linear pricing rule for which unconditional expected market maker profits are zero. Such a pricing rule would require commitment by market makers, because it is not consistent with ex post optimization by market makers.

[7] While the OWR model uses both prices and order flows for estimation, their model shares the feature of the PIN model that the unconditional order flow distribution depends on the information asymmetry parameters and hence could be used to identify information asymmetry. This is inconsistent with our theoretical result that both prices and order flows are necessary to identify alpha when a strategic trader trades endogenously.

trading is quite positively correlated across models, resulting in the positive correlation of the composite measures. Of course, applications of the measures generally assume that they are correlated with private information, not just inversely correlated with liquidity trading.

Theory predicts that orders have larger price impacts and quoted spreads when information asymmetry is more severe.[8] This is true in both the Kyle (1985) model, on which the hybrid and OWR models are based, and the Glosten and Milgrom (1985) model, on which PIN models are based. To test this implication of theory, we examine reduced-form price impacts for our sample as well as quoted spreads. Empirically, expected average lambda from the hybrid model is positively correlated with price impacts and quoted spreads both in the time series and cross-sectionally. While the same is also true for PIN, APIN, VPIN, and the OWR lambda, expected average lambda has a higher correlation with price impacts and spreads in the time series than do the other composite measures. Expected average lambda also adds explanatory power relative to the other measures in cross-sectional regressions of price impacts or quoted spreads on the composite measures.

Other related theoretical work includes Rossi and Tinn (2010), Foster and Viswanathan (1995), Chakraborty and Yilmaz (2004), Goldstein and Guembel (2008), Banerjee and Breon-Drish (2017), and Wang and Yang (2017). Rossi and Tinn solve a two-period Kyle model in which there are two large traders, one of whom is certainly informed and one of whom may or may not be informed. In their model, unlike ours, there are always information events. Foster and Viswanathan (1995) consider a series of single-period Kyle models in which traders choose in each period whether to pay a fee to become informed. There may be periods in which there are no informed traders. However, in their model, it is always common knowledge how many traders choose to become informed, so, in contrast to our model, there is no learning from orders about whether informed traders are present.

Chakraborty and Yilmaz (2004) and Goldstein and Guembel (2008) study discrete-time Kyle models in which there may or may not be an information event. The main result in Chakraborty and Yilmaz (2004) is that the informed trader will manipulate (sometimes buying when she has bad information and/or selling when she has good information) if the horizon is sufficiently long. The primary difference between their model and ours is that they assume that the liquidity trade distribution has finite support, so market makers may incorrectly rule out a type of trader if the horizon is sufficiently long. In contrast, market makers in our model can never rule out any type of the informed trader until the end of the model, so it does not strictly pay for a low type

---

[8] There seems to be general agreement that at least a portion of the price impact of trades is due to information asymmetry. Glosten and Harris (1988), Hasbrouck (1988), and Hasbrouck (1991) estimate models of trades and price changes in which both information asymmetry and inventory control motives are accommodated, and all three papers conclude that information asymmetry is important.

to pretend to be a high type or vice versa. The primary focus of Goldstein and Guembel (2008) concerns the incentives for an uninformed strategic trader to manipulate if information in financial markets feeds back into managers' investment decisions. In their benchmark equilibrium with no feedback, the uninformed speculator behaves as a contrarian but does not manipulate, which is the case in our equilibrium.

Banerjee and Breon-Drish (2017) and Wang and Yang (2017) study continuous-time Kyle models (specifically, the model of Back and Baruch (2004) in which there is a random announcement date) in which an informed trader may not be present. Banerjee and Breon-Drish study the information acquisition decision, treating it as a real option. In one version of their model, the timing of information acquisition is publicly observed. In that version, the market is infinitely deep before information is acquired, and the model is essentially the same as in Back and Baruch after information is acquired. In a second version of their model, the timing of information acquisition is not publicly observed, and the market tries to learn from orders whether information has been acquired. For that version, they establish a nonexistence result: In the class of pricing rules they consider, there is no equilibrium. Wang and Yang also study the Back-Baruch version of the Kyle model. In their model, nature chooses at date 0 whether there is an information event (and all information events are "good news" events). Unlike in our model or the model of Banerjee and Breon-Drish, the strategic trader is not present in their model when there is no information event.[9] They also show the nonexistence of equilibria (though they have an existence result for a second version of their model in which the market maker is a monopolist).

1. The Hybrid Model

The hybrid model includes two important features of PIN models—a probability less than 1 of an information event and a binary asset value conditional on an information event—and it also includes an optimizing (possibly) informed trader, like in the Kyle (1985) model. Denote the time horizon for trading by $[0, 1]$. Assume there is a single risk-neutral strategic trader. Assume this trader receives a signal $S \in \{L, H\}$ at time 0 with probability α, where $L < 0 < H$.[10] Let p_L and $p_H = 1 - p_L$ denote the probabilities of low and high signals, respectively, conditional on an information event. With probability $1 - \alpha$, there is no information event, and the trader also knows when this happens. Let ξ denote an indicator for whether an information event has occurred ($\xi = 1$ if yes and $\xi = 0$ if no). In addition to the private information, public information can also arrive during the course of trading, represented

---

[9] We call the strategic trader when there is no information event a "contrarian trader." See Section 1.2 for a discussion.

[10] Internet Appendix A extends the model to general signal distributions.

by a martingale V. The possible private information—whether there was an information event and, if so, whether the signal was low or high—becomes public information after the close of trading at date 1, producing an asset value of $V_1 + \xi S$. Without loss of generality, we take the signal S to have a zero mean. We can always do this by taking the signal mean to be part of the public information V_0.

In addition to the strategic trades, there are liquidity trades represented by a Brownian motion Z with zero drift and instantaneous standard deviation σ. Let X_t denote the number of shares held by the strategic trader at date t (taking $X_0 = 0$ without loss of generality), and set $Y_t = X_t + Z_t$. The processes Y and V are observed by market makers. Denote the information of market makers at date t by $\mathcal{F}_t^{V,Y}$.

One requirement for equilibrium in this model is that the price equal the expected value of the asset conditional on the market makers' information and given the trading strategy of the strategic trader:

$$P_t = \mathsf{E}\big[V_1 + \xi S \,|\, \mathcal{F}_t^{V,Y}\big] = V_t + \mathsf{E}\big[\xi S \,|\, \mathcal{F}_t^{V,Y}\big]. \tag{1}$$

We will show that there is an equilibrium in which $P_t = V_t + p(t, Y_t)$ for a function p. This means that the expected value of ξS conditional on market makers' information depends only on cumulative orders Y_t and not on the entire history of orders.

The other requirement for equilibrium is that the strategic trades are optimal. Let θ_t denote the trading rate of the strategic trader (i.e., $dX_t = \theta_t \, dt$). The process θ has to be adapted to the information possessed by the strategic trader, which is V, ξS, and the history of Z (in equilibrium, the price reveals Z to the informed trader). The strategic trader chooses the rate to maximize

$$\mathsf{E} \int_0^1 [V_1 + \xi S - P_t] \theta_t \, dt = \mathsf{E} \int_0^1 [\xi S - p(t, Y_t)] \theta_t \, dt, \tag{2}$$

with the function p being regarded by the informed trader as exogenous. In the optimization, we assume that the strategic trader is constrained to satisfy the "no doubling strategies" condition introduced in Back (1992), meaning that the strategy must be such that

$$\mathsf{E} \int_0^1 p(t, Y_t)^2 \, dt < \infty$$

with probability 1.

Let N denote the standard normal distribution function, and let n denote the standard normal density function. Set $y_L = \sigma \, \mathrm{N}^{-1}(\alpha p_L)$ and $y_H = \sigma \, \mathrm{N}^{-1}(1 - \alpha p_H)$. This means that the probability mass in the lower tail $(-\infty, y_L)$ of the distribution of cumulative liquidity trades Z_1 equals αp_L, which is the unconditional probability of bad news. Likewise, the probability mass in

the upper tail (y_H, ∞) of the distribution of Z_1 equals αp_H, which is the unconditional probability of good news. Set

$$q(t,y,s) = \begin{cases} \mathsf{E}[Z_1 - Z_t \mid Z_t = y, Z_1 < y_L] & \text{if } s = L, \\ \mathsf{E}[Z_1 - Z_t \mid Z_t = y, y_L \leq Z_1 \leq y_H] & \text{if } s = 0, \\ \mathsf{E}[Z_1 - Z_t \mid Z_t = y, Z_1 > y_H] & \text{if } s = H. \end{cases} \quad (3)$$

From the standard formula for the mean of a truncated normal, we obtain the following more explicit formula for q:

$$\frac{q(t,y,s)}{\sigma\sqrt{1-t}} = \begin{cases} -\mathrm{n}\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) \Big/ \mathrm{N}\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) & \text{if } s = L, \\ \left[\mathrm{n}\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) - \mathrm{n}\left(\frac{y_H - y}{\sigma\sqrt{1-t}}\right)\right] \Big/ \left[\mathrm{N}\left(\frac{y_H - y}{\sigma\sqrt{1-t}}\right) - \mathrm{N}\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right)\right] & \\ & \text{if } s = 0, \\ \mathrm{n}\left(\frac{y - y_H}{\sigma\sqrt{1-t}}\right) \Big/ \mathrm{N}\left(\frac{y - y_H}{\sigma\sqrt{1-t}}\right) & \text{if } s = H. \end{cases} \quad (4)$$

The equilibrium described in Theorem 1 below can be shown to be the unique equilibrium in a certain broad class, following Back (1992). The proof of Theorem 1 is given in Appendix A.[11]

Theorem 1. There is an equilibrium in which the trading rate of the strategic trader is

$$\theta_t = \frac{q(t, Y_t, \xi S)}{1 - t}. \quad (5)$$

Given market makers' information at any date t, the conditional probability of an information event with a low signal is $\mathrm{N}\left(\frac{y_L - Y_t}{\sigma\sqrt{1-t}}\right)$ and the conditional probability of an information event with a high signal is $\mathrm{N}\left(\frac{Y_t - y_H}{\sigma\sqrt{1-t}}\right)$. The equilibrium asset price is $P_t = V_t + p(t, Y_t)$, where the pricing function p is given by

$$p(t,y) = L \cdot \mathrm{N}\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) + H \cdot \mathrm{N}\left(\frac{y - y_H}{\sigma\sqrt{1-t}}\right). \quad (6)$$

In this equilibrium, the process Y is a martingale given market makers' information and has the same unconditional distribution as does the liquidity trade process Z; that is, it is a Brownian motion with zero drift and standard deviation σ.

The last statement of the theorem implies that the distribution of order flows in the model does not depend on the information asymmetry parameters α,

---

[11] The proof is based on a generalization of the Brownian bridge feature of the continuous-time Kyle model established in Back (1992). Whereas a Brownian bridge is a Brownian motion conditioned to end at a particular point, in this model (with a discrete rather than continuous distribution of the asset value) we encounter a Brownian motion conditioned only to end in a particular interval. The generalization of the Brownian bridge is established as a lemma in Appendix A.
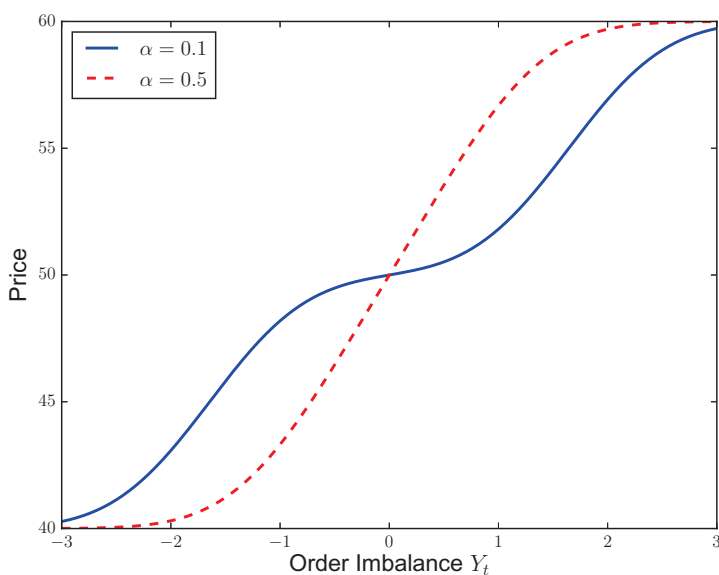
Figure 1
The equilibrium price $V_t + p(t, Y_t)$ as a function of the order imbalance Y_t
The parameter values are $t = 0.5$, $V_t = 50$, $H = 10$, $L = -10$, $\sigma = 1$, and $p_H = p_L = 1/2$.

H, and L. Thus, if the model is correct, it is impossible to estimate those parameters using order flows alone. In general, the theorem suggests that it may be difficult to identify information asymmetry parameters using order flows alone, as discussed in the Introduction and Section 1.1. When we estimate the hybrid model, we use both order flows and returns, in contrast to related models that only use order flows.

Empirically, we test the relationship between α and price impacts of trades. Figure 1 plots the equilibrium price as a function of Y_t for two different values of α. It shows that the price is more sensitive to orders when α is larger. To investigate further how the sensitivity of prices to orders depends on α in the hybrid model, we calculate the price sensitivity—that is, we calculate Kyle's lambda.

Theorem 2. In the equilibrium of Theorem 1, the asset price evolves as $dP_t = dV_t + \lambda(t, Y_t)dY_t$, where Kyle's lambda is

$$\lambda(t, y) = -\frac{L}{\sigma\sqrt{1-t}} \cdot n\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) + \frac{H}{\sigma\sqrt{1-t}} \cdot n\left(\frac{y_H - y}{\sigma\sqrt{1-t}}\right). \quad (7)$$

Furthermore, Kyle's lambda $\lambda(t, Y_t)$ is a martingale with respect to market makers' information on the time interval $[0, 1)$.

Kyle's lambda is a stochastic process in our model, but we can easily relate the expected average lambda to α. Because lambda is a martingale, the expected
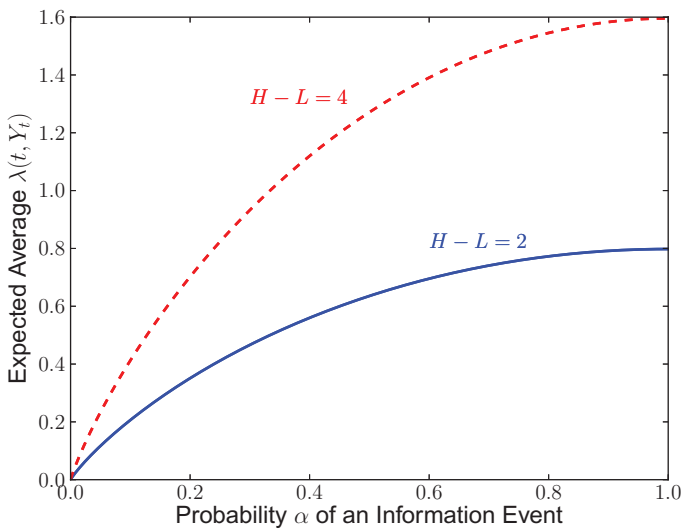
Figure 2
Expected average lambda (8) **as a function of** α
The parameter values are $\sigma = 1$, $p_L = p_H = 1/2$, and $L = -H$.

average lambda is $\lambda(0,0)$. Substitute the definitions of y_L and y_H in (7) to compute[12]

$$\lambda(0,0) = -\frac{L}{\sigma} \mathrm{n}\left(\mathrm{N}^{-1}(\alpha p_L)\right) + \frac{H}{\sigma} \mathrm{n}\left(\mathrm{N}^{-1}(1 - \alpha p_H)\right). \tag{8}$$

Figure 2 plots the expected average lambda as a function of α for two values of H, taking $L = -H$. Doubling the signal magnitudes doubles lambda. Furthermore, the expected average lambda is increasing in α.

1.1 Nonidentifiability using order flows alone
A key result of Theorem 1 is that the aggregate order imbalance Y_1 has the same distribution as the liquidity trades Z_1 and is invariant with respect to the information asymmetry parameters.[13] Further insight into this identification issue can be gained by noting that the unconditional distribution of the order imbalance in our model is a mixture of three conditional distributions. With

---

[12] If information events occur for sure ($\alpha = 1$), then $\lambda(0,0) = (H - L)\mathrm{n}(0)/\sigma$. This is analogous to the result of Kyle (1985) that lambda is the ratio of the signal standard deviation to the standard deviation of liquidity trading. Of course, it is not quite the same as Kyle's formula, because we have a binary signal distribution, whereas the distribution is normal in Kyle (1985).

[13] This result on the nonidentifiability of information asymmetry parameters from order flows does not depend on the binary signal assumption. Internet Appendix A presents the model with a general signal distribution. The unconditional order flow distribution is the same as the distribution of liquidity order flows in the general model as well.
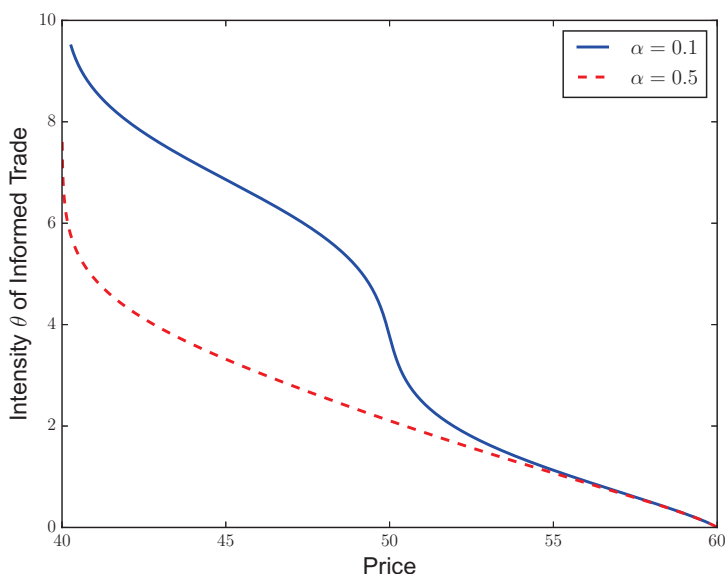
Figure 3
The equilibrium informed trading rate θ_t as a function of the price $V_t + p(t, Y_t)$
The parameter values are $t = 0.5$, $\xi S = H$, $V_t = 50$, $H = 10$, $L = -10$, $\sigma = 1$, and $p_H = p_L = 1/2$.

probability αp_L, Y_1 is drawn from the distribution conditional on a low signal; with probability αp_H, Y_1 is drawn from the distribution conditional on a high signal; and with probability $1 - \alpha$, Y_1 is drawn from the distribution conditional on no information event. The first two distributions have nonzero means—there is an excess of sells over buys in the first and an excess of buys over sells in the second. One might conjecture that changing α—thereby changing the likelihood of drawing from the first two distributions—will alter the unconditional distribution of Y_1. If so, then one could perhaps identify α from the distribution of Y_1. In other models with a potential information event, it is indeed true that changing α, holding other parameters constant, alters the unconditional distribution of the order imbalance. However, it is not true in our model, because the distribution of informed trades in our model endogenously depends on α due to liquidity depending on α. With a larger alpha, the market is less liquid (see the comparative statics in Figure 2) and the informed trader trades less aggressively. Furthermore, with endogenous informed orders, the arrival rate of informed orders depends on prior price changes as shown in Figure 3, which is not the case in other models with a potential information event. In particular, when prices have moved in the direction of the news, informed orders slow down, and, when prices have moved in the opposite direction, informed orders speed up. Figure 3 shows that these changes in intensity depend on the ex ante probability α of an information event. Thus, the
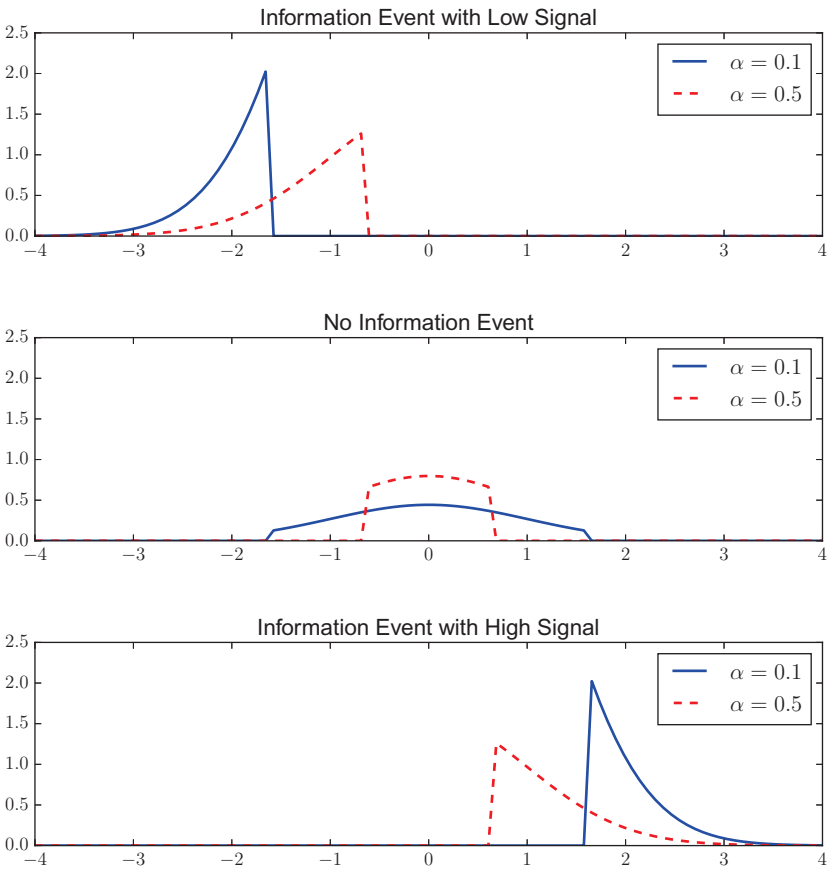
Figure 4
The conditional density function of the net order flow Y_1
The density is conditional on a low signal, no information event, or a high signal. The parameter values are $\sigma = 1$ and $p_L = p_H = 1/2$.

distributions over which we are mixing change when the mixture probabilities change, leaving the unconditional distribution of Y_1 invariant with respect to α.

The change in the conditional distributions is illustrated in Figure 4. The top and bottom panels of Figure 4 show that the strategic trader trades more aggressively when an information event occurs if an information event is less likely ($\alpha = 0.1$ versus $\alpha = 0.5$). The unconditional distribution of Y_1 is standard normal for both $\alpha = 0.1$ and $\alpha = 0.5$ in Figure 4, so we cannot hope to use the unconditional distribution to recover α.

Of course, identifying the information asymmetry parameters from the distribution of order imbalances is a very different issue from using order imbalances to update the probability of an information event in a particular

instance of the model. Conditional on knowledge of the parameters, the order imbalance does help in estimating whether an information event occurred in a particular instance of the model; in fact, the market makers in the model update their beliefs regarding the occurrence of an information event based on the order imbalance. So, we can compute

$$\text{prob}(\text{info event} \mid Y_t, \text{parameters}),$$

and this probability does depend on the information asymmetry parameters. We could use this to identify the information asymmetry parameters if we had data on order imbalances *and data on whether information events occurred.* Of course, we generally do not have data of the latter type. Theorem 1 shows that the likelihood function of the information asymmetry parameters given only data on order imbalances is a constant function of those parameters; hence, the order imbalances alone cannot identify them.

In our empirical work, we estimate the model parameters using prices and order flows. Armed with these parameter estimates and order flow observations, we can compute conditional probabilities of an information event. We examine their time-series properties around earnings announcements and around Schedule 13D filer trades in Section 3.1.

1.2 The contrarian trader assumption

One way in which our model departs from related models like the PIN model is that the strategic trader is present in our model even when there is no information event. When there is no information event, this trader behaves as a contrarian, selling on price increases and buying on price declines.[14] The existence of such a contrarian trader seems likely if there are always some traders who are best informed—corporate managers, for example. This would be the case if information were truly idiosyncratic to the firm. If, on the other hand, there is an industry or other aggregate components to the information, then it is possible that no one knows when no one else has information. In that case, the contrarian trader that we posit would not exist.

In Internet Appendix B, we solve a variant of the PIN model in which contrarian traders arrive at the market when there is no information event. The contrarian traders condition their trading direction on the prevailing bid and ask quotes and the intrinsic value of the asset. The distribution of order imbalances in that model is shown in Figure 5 for three different values of α (the probability of an information event). The figure shows that the distribution depends on α; thus, order imbalances can be used to identify information asymmetry in

---

[14] We assume the existence of such a trader because it makes the model more tractable. Odders-White and Ready (2008) describe the trader as also being present in their model when there is no information event, but, because the trader has no opportunity to react to price changes in their one-period model, the trader optimally chooses a zero trade in the absence of an information event. Goldstein and Guembel (2008) also assume that the uninformed speculator trades as a contrarian in their benchmark model with no feedback.

(a) $\alpha = 0.25$

(b) $\alpha = 0.50$

(c) $\alpha = 0.75$
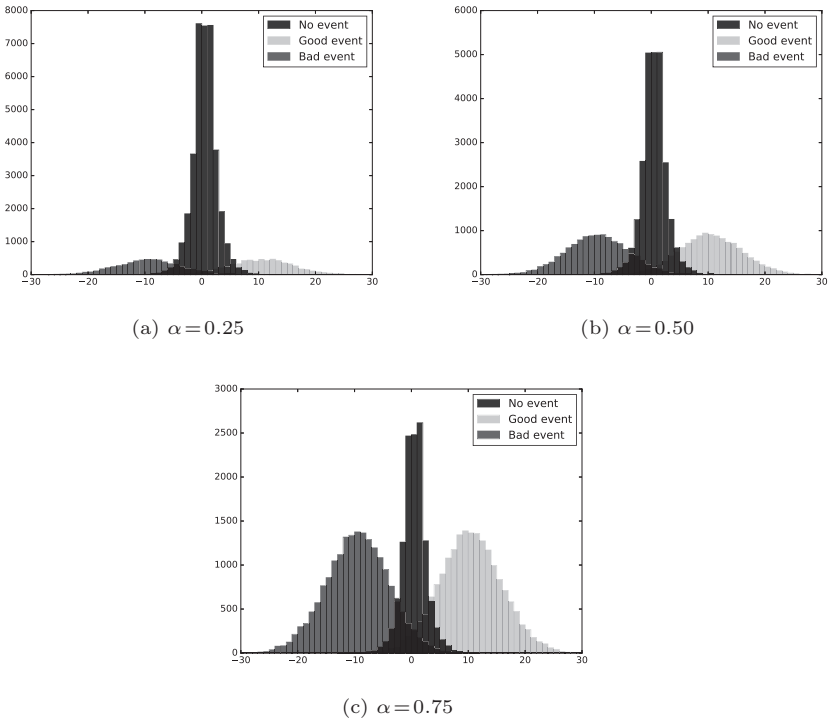
Figure 5
The simulated distribution of order imbalances for a variant of the Easley et al. (1996) model in which contrarian traders arrive in the event of no information
The model is described in Internet Appendix B. Order imbalance is the number of buys minus number of sells. The histograms plot 50,000 instances of the model. The parameter values are $\alpha \in \{0.25, 0.5, 0.75\}$, $p_L = 0.5$, $\varepsilon = 10$, $\mu = 10$, $L = -1$, $H = 1$, and $V^* = 0$.

the PIN model even when a contrarian trader is present. Thus, the contrarian trader assumption is not the main driving force behind our nonidentifiability result. Instead, the result depends on market makers reacting to information asymmetry and on strategic traders reacting both to liquidity and to price changes. That is, order flows depend on market liquidity, which depends on information asymmetry. This creates an indirect dependence of order flows on information asymmetry that is countervailing to the direct relation.

2. Estimation of the Model

We estimate the hybrid model using trade and quote data from TAQ for NYSE firms from 1993 through 2012.[15] We sign trades as buys and sells using the

___

[15] We require that firms have intraday trading observations for at least 200 days within the year. We also require firms have the same ticker throughout the year and experience no stock splits.

Lee and Ready (1991) algorithm: trades above (below) the prevailing quote midpoint are considered buys (sells). If a trade occurs at the midpoint, then the trade is classified as a buy (sell) if the trade price is greater (less) than the previous differing transaction price.[16] We sample prices and order imbalances hourly and at the close and define order imbalances as shares bought less shares sold (denoted in thousands of shares).

We estimate the model by maximum likelihood, maintaining the standard assumptions in the literature that each day is a separate realization of the model and that parameters are constant within each year for each stock. We assume that the dispersion of the possible signals on each day i is proportional to the observed opening price on day i, P_{i0}. Specifically, we assume that, for each firm-year, there is a parameter κ such that the low signal value each day is $L = -2 p_H \kappa P_{i0}$ and the high signal value is $H = 2 p_L \kappa P_{i0}$. This construction ensures that the signal has a zero mean and $(H - L)/P_{i0} = 2\kappa$. Thus, κ measures the signal magnitude. We also assume that the public information process V is a geometric Brownian motion on each day with a constant volatility δ. The likelihood function for the hybrid model depends on the signal magnitude κ, the probability α of information events, the probability p_L of a negative signal conditional on an information event, the standard deviation σ of liquidity trading, and the volatility δ of public information.

We derive the likelihood function for the model in Appendix B. Dropping constants, the log-likelihood function $\mathcal{L}$ for an observation period of n days satisfies

$$-\mathcal{L} = n(k+1)\log\sigma + \frac{1}{2\sigma^2\Delta}\sum_{i=1}^{n} Y_i' \Sigma^{-1} Y_i + n(k+1)\log\delta$$

$$+ \frac{1}{2\delta^2\Delta}\sum_{i=1}^{n} U_i' \Sigma^{-1} U_i + \frac{n\delta^2}{8} + \sum_{i=1}^{n}\left(\sum_{j=1}^{k} U_{ij} + \frac{3}{2}U_{i,k+1}\right), \qquad (9)$$

where k is the number of intraday observations sampled at regular intervals of length Δ. We sample every hour and at the close, so $k = 6$ and $\Delta = 1/6.5$. Y_i is the vector of cumulative order flows for day i. U_i is the vector $(U_{i1}, \ldots, U_{i,k+1})'$ of log pricing differences

$$U_{ij} = \log\left(\frac{P_{ij}}{P_{i0}} - p(t_j, Y_{ij})\right) \qquad (10)$$

between the observed return and the model's pricing function. Σ is a $(k+1) \times (k+1)$ matrix that depends on Δ as described in Appendix B. We minimize (9) in α, κ, p_L, σ, and δ.

---

[16] Prior to 2000, quotes are lagged 5 seconds when matched to trades. For the 2000–2006 time period, quotes are lagged 1 second. From 2007 on, quotes are matched to trades in the same second.

The private information parameters α, κ, and p_L enter the likelihood function via the log pricing errors U_i, because the parameters affect the pricing function $p(t, Y_t)$. As can be seen from (9), α, κ, and p_L are estimated by minimizing a quadratic function of the log pricing errors. In the model, the pricing errors are due to public information. In minimizing the quadratic function, the estimation procedure tries to maximize the fit of the model prices $p(t_j, Y_{ij})$ to the observed returns and thereby to minimize how much we have to rely on public information to explain the returns.

Figure 6 illustrates how the pricing errors depend on the private information parameters. For simplicity, Figure 6 treats the case $k=0$; that is, it only uses daily order imbalances and returns. The pricing error each day is the difference between the daily return P_1/P_0 and the model price $p(1, Y_1)$. The price function $p(1, \cdot)$ is a step function,[17] with steps at y_L and y_H defined in Section 1 as $y_L = \sigma \, N^{-1}(\alpha p_L)$ and $y_H = \sigma \, N^{-1}(1 - \alpha p_H)$. Thus, α and p_L affect the step locations. If α is larger, the step locations are closer together. If p_L is increased, both step locations shift to the right. The parameter κ determines the height of the steps. Notice that σ and α play similar roles in determining the step locations; either increasing σ or decreasing α will spread out the steps. However, maximizing the likelihood function also involves fitting the order imbalances to a Brownian motion with standard deviation σ. Table 2 (see Section 2.1) shows that our empirical estimates of σ are almost entirely determined by the standard deviations of order imbalances—likewise, the estimates of δ (the standard deviation of the public information process) are almost entirely determined by the standard deviations of returns.

Figure 6 depicts simulated data and three different sets of possible estimates for the parameters α and κ. The fit of the price function $p(1, Y_1)$ to the daily returns is shown in the left column. The log pricing errors in all three cases are shown in the right column. The parameters that were used in the simulation are shown in the middle row. Of the three sets of parameters shown in the figure, the parameters in the middle row give the largest value for the likelihood function. The parameters in the top row produce steps that are too far apart and too small, generating a price function that is too flat compared to the data. Consequently, the log pricing errors shown in the top row of the right column are positively correlated with order imbalances. The parameters in the bottom row produce steps that are too close together and too large, generating a price function that is too steep compared to the data. Consequently, the log pricing errors in the bottom row are negatively correlated with order imbalances.

2.1 Estimates of the hybrid model
Table 1 reports summary statistics of the parameter estimates for the panel of firm-years (summary statistics by year are plotted in Figure 7 in Section 2.5). To

---

[17] The price function $p(t, \cdot)$ for $t < 1$ (that is, for intra-day returns) is depicted in Figure 1.

(a) $\hat{\alpha} = 0.25, \hat{\kappa} = 0.01, \hat{p}_L = 0.5$

(b) $\hat{\alpha} = 0.25, \hat{\kappa} = 0.01, \hat{p}_L = 0.5$

(c) $\hat{\alpha} = 0.50, \hat{\kappa} = 0.015, \hat{p}_L = 0.5$

(d) $\hat{\alpha} = 0.50, \hat{\kappa} = 0.015, \hat{p}_L = 0.5$

(e) $\hat{\alpha} = 0.75, \hat{\kappa} = 0.02, \hat{p}_L = 0.5$

(f) $\hat{\alpha} = 0.75, \hat{\kappa} = 0.02, \hat{p}_L = 0.5$
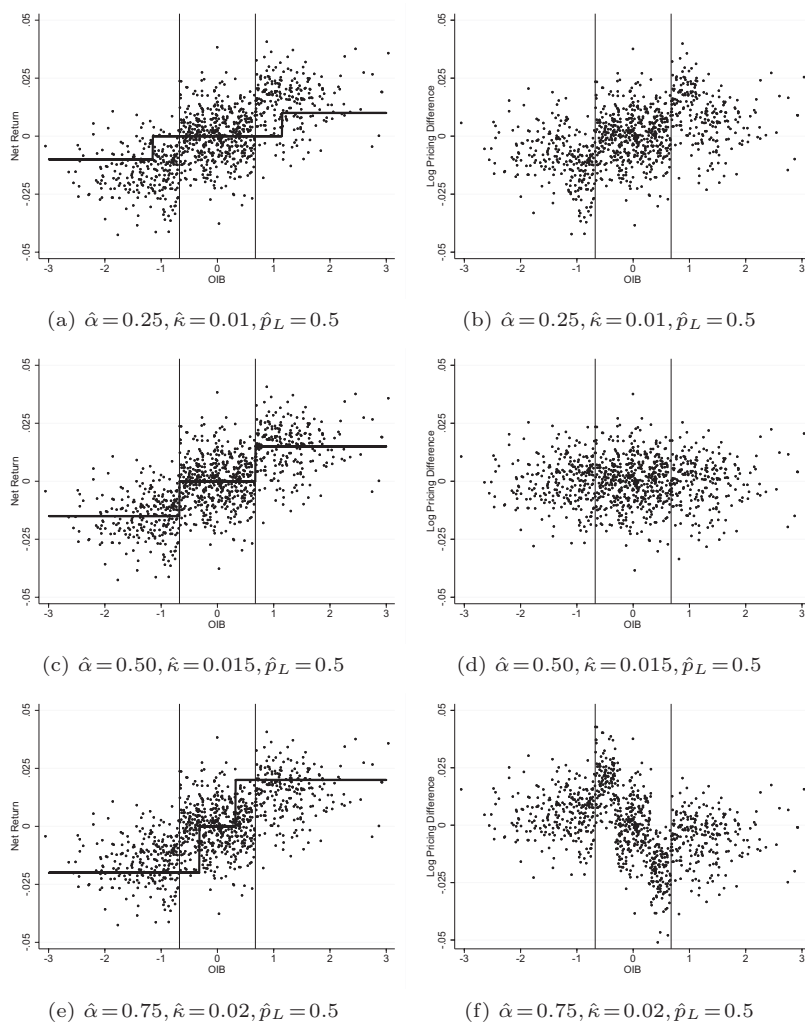
Figure 6
Returns, order flows, and log pricing differences for various parameters
Simulations of 1,000 instances of the hybrid model. The data-generating parameters are $\alpha = 0.5$, $\kappa = 0.015$, $p_L = 0.5$, $\sigma = 0.1$, $\delta = 0.01$. Standardized order flows are on the horizontal axis. The left column plots end-of-day net returns, $P_1/P_0 - 1$, and the pricing function, $p(1, Y_1)$. The right column plots log pricing differences, $U_1 = \ln(P_1/P_0 - p(1, Y_1))$. The pricing function $p(1, Y_1)$ depends on the indicated hatted parameters in each panel caption. Each row plots the pricing function and log pricing differences for different parameter estimates (hatted values). The vertical lines indicate the thresholds y_L/σ and y_H/σ for the true parameters. The first row uses parameter estimates in which α and κ are too low relative to the true parameters. These generate log pricing differences that are positively correlated with order flows. The second row uses the data-generating parameters. The log pricing differences are uncorrelated with order flows. The third row uses parameter estimates in which α and κ are too high relative to the true parameters. These generate log pricing differences that are negatively correlated with order flows.

Table 1
Hybrid model parameter estimate summary statistics

	α	κ	p_L	σ	δ
Mean	0.64	0.0068	0.51	0.12	0.0213
SD	0.25	0.0050	0.15	0.11	0.0087
First quartile	0.54	0.0032	0.46	0.05	0.0149
Median	0.68	0.0058	0.50	0.08	0.0197
Third quartile	0.81	0.0095	0.56	0.16	0.0258
N	19,965	19,965	19,965	19,965	19,965

The model is estimated on a stock-year basis for NYSE stocks from 1993 to 2012 using prices and order imbalances in 6 hourly intraday bins and at the close. The model parameters are α = probability of an information event, κ = signal scale parameter, σ = standard deviation of liquidity trading, δ = volatility of public information, and p_L = probability of a negative event.

see which aspects of the data determine the parameter estimates, Table 2 reports regressions of the parameter estimates on various moments of order flows and returns. The table also reports variance decompositions. The moments include correlations of order flows and returns split into two subperiods of the day: the first 3 hours and the last 3.5 hours. The price function in the model is nonlinear, so we also include nonlinear measures of the comovement of returns and order imbalances. Specifically, we include correlations of returns with squared order imbalances for the two subperiods. We also include the fraction of the days on which returns and order imbalances are both in the right tails of their distributions and the fraction in which they are both in their left tails, defining a tail as a standard deviation away from zero (a zero order imbalance or a zero rate of return).

The R-squareds and the variance decomposition show that the estimates of the standard deviation σ of order imbalances from the model are almost entirely determined by the empirical standard deviations of order imbalances. Likewise, the estimates of the volatility δ of the public news process are almost entirely determined by the standard deviations of returns. The private information parameters κ, α, and p_L are naturally more complex.

The moments have little explanatory power for the p_L estimates. As shown in Table 1, the distribution of the p_L estimates is fairly tight around 50%, so there is not too much variation to explain. The κ and α estimates are the most interesting. The magnitude κ of private information is fairly well explained by the moments, with the most important moments being the standard deviation of returns and the correlations between order imbalances and returns. The variance decomposition shows that all of the moments except skewness affect the estimated probability α of information events. The nonlinear specification is important for α. More than 20% of the R-squared comes from the tail variables.

2.2 Testing whether an information event is always present in the hybrid model
Our hybrid model relaxes the assumption in Kyle (1985) that an information event occurs in each instance of the model (in each day in our implementation).

Table 2
Hybrid model parameter estima tes and moments of order flow and returns

A. Standardized Regression

	α	κ	p_L	σ	δ
sd(OIB)	−0.129***	0.007	−0.089***	0.986***	−0.000
	(−5.57)	(0.38)	(−6.17)	(135.67)	(−0.02)
sd(R)	0.155***	0.460***	0.016	−0.007	0.963***
	(5.15)	(7.89)	(1.39)	(−1.46)	(138.47)
skew(OIB)	0.007	0.003	−0.058***	0.003	0.006*
	(1.02)	(0.39)	(−6.11)	(0.79)	(1.69)
skew(R)	−0.008	0.009	0.047***	−0.001	0.005*
	(−1.05)	(1.51)	(4.33)	(−0.41)	(1.95)
corr(R_1, OIB$_1$)	0.258***	0.484***	−0.018	0.009	0.039***
	(5.40)	(17.25)	(−0.80)	(1.26)	(2.96)
corr(R_1, OIB$_1^2$)	−0.039***	−0.018	0.185***	−0.003	−0.008*
	(−3.16)	(−1.29)	(5.73)	(−1.12)	(−1.92)
corr(R_2, OIB$_2$)	0.218***	0.314***	−0.034	−0.012**	−0.022**
	(6.10)	(14.92)	(−1.26)	(−2.14)	(−1.97)
corr(R_2, OIB$_2^2$)	−0.049***	−0.028**	0.099***	−0.001	−0.009**
	(−5.79)	(−2.04)	(4.19)	(−0.41)	(−2.52)
# right tail OIB & R	−0.122***	−0.103***	−0.128***	0.011*	−0.074***
	(−4.17)	(−5.59)	(−3.86)	(1.76)	(−5.95)
# left tail OIB & R	−0.163***	−0.063***	0.029	0.005	0.012*
	(−7.39)	(−6.66)	(1.38)	(0.65)	(1.67)
Constant	2.159***	−0.482***	3.439***	0.068***	0.118***
	(17.04)	(−4.53)	(60.66)	(3.56)	(5.39)
Observations	19,965	19,965	19,965	19,965	19,965
Adjusted R^2	0.152	0.680	0.040	0.978	0.938

B. Variance Decomposition

	α	κ	p_L	σ	δ
sd(OIB)	0.125	0.000	0.127	1.000	0.000
sd(R)	0.237	0.636	0.005	0.000	0.997
skew(OIB)	0.000	0.000	0.075	0.000	0.000
skew(R)	0.001	0.000	0.047	0.000	0.000
corr(R_1, OIB$_1$)	0.221	0.240	0.002	0.000	0.001
corr(R_1, OIB$_1^2$)	0.009	0.001	0.458	0.000	0.000
corr(R_2, OIB$_2$)	0.159	0.101	0.008	0.000	0.000
corr(R_2, OIB$_2^2$)	0.016	0.002	0.137	0.000	0.000
# right tail OIB & R	0.055	0.012	0.128	0.000	0.002
# left tail OIB & R	0.176	0.008	0.012	0.000	0.000
Observations	19,965	19,965	19,965	19,965	19,965
Adjusted R^2	0.152	0.680	0.040	0.978	0.938

The dependent variables are the estimated parameters from the hybrid model. The explanatory variables are various moments of order flows and returns. The unit of observation is a firm-year. OIB denotes the cumulative order flow over the full day. OIB$_1$ and OIB$_2$ are the order flows over the first 3 and last 3.5 hours of the trading day. Similarly, R is the return over the full day, and R_1 and R_2 are returns over the first 3 and last 3.5 hours of the trading day. The indicated moments of these variables are calculated across days for each firm-year. # Right Tail OIB & R is the fraction of days where both OIB > sd(OIB) and $R − 1$ > sd(R). # Left Tail OIB & R is the fraction of days where both OIB < −sd(OIB) and $R − 1$ < −sd(R). Panel A reports estimates where all variables are standardized to have a unit standard deviation. Standard errors are clustered by firm and year. t-statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Panel B reports variance decompositions. Each number in panel B represents the fraction of the model's total partial sum of squares corresponding to the moment in the row. The sum of each column is thus one.

A natural question is whether this relaxation is supported in the data. The Kyle framework is nested in our model by the restriction that $\alpha = 1$. Accordingly, we estimate the model with this restriction. The standard likelihood ratio test of the null that $\alpha = 1$ against the alternative that $\alpha \in [0, 1]$ is rejected for 73% of the firm-years (with a test size of 10%). However, the usual regularity conditions for the likelihood ratio test require that the restriction not be at the boundary of the parameter space. To address this issue, we bootstrap the distribution of the likelihood ratio statistic for a random sample of 100 firm-years like in Duarte and Young (2009).

Specifically, for a given firm-year, we estimate the restricted model ($\alpha = 1$) and then simulate 500 firm-years under the null using the estimated (restricted) parameters. We then estimate the restricted and unrestricted models for each simulated firm-year to obtain the distribution of the likelihood ratio under the null. The 90th percentile of this distribution is the critical value to evaluate the empirical likelihood ratio. These bootstrapped likelihood ratio tests reject the restricted Kyle model in favor of the hybrid model for 62 of the 100 randomly selected firm-years. The data thus supports the conclusion that the probability of an information event is less than 1.

2.3 Estimated parameters and reduced-form price impacts

The model places structure on the price and order flow data, allowing the econometrician to identify components of Kyle's lambda. Of course, one can estimate a reduced-form price impact as well. As an initial test of whether our estimates relate to price impact as implied by theory, we test the comparative statics from Figure 2 that price impacts are increasing in both the probability and magnitude of information events.

We employ three estimates of the price impact of orders. The first is the 5-minute percent price impact of a given trade k as

$$5\text{-}minute\ price\ impact_k = \frac{2D_k(M_{k+5} - M_k)}{M_k}, \tag{11}$$

where M_k is the prevailing quote midpoint for trade k, M_{k+5} is the quote midpoint five minutes after trade k, and D_k equals 1 if trade k is a buy and -1 if trade k is a sell. Goyenko, Holden, and Trzcinka (2009) use this measure as one of their high-frequency liquidity benchmarks in a study assessing the quality of various liquidity measures based on daily data.[18] For a given stock-day, the estimate of the percent price impact is the equal-weighted average price impact over all trades on that day. We average these daily price impact estimates for each stock-year.

---

[18] Holden and Jacobsen (2014) show that liquidity measures such as the percent price impact can be biased when constructed from monthly TAQ data, so we follow their suggested technique in processing the data.

Table 3
Panel regressions of price impacts

A. Probability and magnitude of information events

	5-minute price impact		Cumulative impulse response		$\widehat{\lambda}_{\text{intraday}}$	
	(1)	(2)	(3)	(4)	(5)	(6)
α	0.22***	0.09***	0.17***	0.06***	0.23***	0.12***
	(5.15)	(4.00)	(3.95)	(2.93)	(4.88)	(4.13)
κ	0.58***	0.35***	0.42***	0.23***	0.67***	0.48***
	(16.03)	(9.29)	(9.86)	(6.44)	(10.74)	(8.27)
Observations	19,965	19,965	19,965	19,965	19,965	19,965
R^2	0.591	0.800	0.625	0.829	0.369	0.642
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	No	Yes	No	Yes	No	Yes

B. Unconditional signal standard deviation

	5-minute price impact		Cumulative impulse response		$\widehat{\lambda}_{\text{intraday}}$	
	(1)	(2)	(3)	(4)	(5)	(6)
SD(ξS)	0.72***	0.50***	0.54***	0.34***	0.83***	0.67***
	(26.04)	(18.11)	(13.27)	(8.72)	(11.64)	(12.46)
Observations	19,965	19,965	19,965	19,965	19,965	19,965
R^2	0.635	0.823	0.655	0.842	0.438	0.679
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	No	Yes	No	Yes	No	Yes

The independent variables are the estimated probability α of an information event, the magnitude κ of an information event (panel A) and the standard deviation of the signal (SD(ξS)) (panel B). The dependent variables are the 5-minute price impact, the cumulative impulse response estimated following Hasbrouck (1991), and an estimate of price impact ($\widehat{\lambda}_{\text{intraday}}$) using a regression of 5-minute returns on the square root of signed volume following Hasbrouck (2009) and Goyenko, Holden, and Trzcinka (2009). All variables are standardized to have a unit standard deviation. Standard errors are clustered by firm and year. *t*-statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

We also estimate the cumulative impulse response function (Hasbrouck, 1991), which captures the permanent price impact of an order. The cumulative impulse response is calculated from a vector autoregression of log price changes and signed trades. Finally, we estimate another price impact measure (denoted $\widehat{\lambda}_{\text{intraday}}$) using a regression of 5-minute returns on the square root of signed volume following Hasbrouck (2009) and Goyenko, Holden, and Trzcinka (2009). We estimate these for each stock day, taking the median estimate across days as the stock-year estimate.

The first panel of Table 3 reports panel regressions of the three price impact measures on the hybrid model parameters that measure private information (the probability α of an information event and the magnitude κ of information events). Before running the regressions, the price impacts and the structural parameters are winsorized at 1% and 99% and standardized to have unit standard deviations. Price impacts are positively related to both α and κ. The coefficients are positive even with the inclusion of firm fixed effects, indicating that α and κ capture within-firm information asymmetry variation as well.

A summary measure of the amount of private information is the standard deviation of the signal ξS, denoted SD(ξS), which equals

$$2\kappa\sqrt{\alpha p_L(1-p_L)}. \tag{12}$$

The second panel of Table 3 shows that the estimated SD(ξS) is strongly positively correlated with the price impact estimates, as expected. Cross-sectionally, a 1-standard-deviation increase in SD(ξS) is associated with around three-quarters of a standard deviation increase in 5-minute price impact and $\widehat{\lambda}_{\text{intraday}}$ and about half a standard deviation increase in the cumulative impulse response measure. Variation in SD(ξS) within firm is positively correlated with within-firm variation in all three price impact measures.

2.4 Kyle's lambda and stochastic volatility

In the model, prices evolve as $dP_t = dV_t + \lambda(t, Y_t)dY_t$. The changing sensitivity of prices to order flows means that prices exhibit stochastic volatility. In Table 4, we investigate this implication of the model for simulated and actual data. Volatility is measured as the absolute return over the last 3.5 hours of the trading day. We calculate $\lambda(t, Y_t)$ from Equation (7) for each day using the cumulative order imbalance over the first 3 hours of the day (i.e., t=3/6.5), along with the estimated parameters. We report predictive regressions of volatility on $\lambda(t, Y_t)$.

The top panel of Table 4 reports results for a simulated panel created by generating 252 days for each set of parameter estimates. Higher levels of $\lambda(t, Y_t)$ predict higher volatility in the second part of the day. The bottom panel shows that this phenomenon holds in the actual data as well. Moreover, the magnitudes are similar across the simulated and actual data controlling for firm and year fixed effects. Confidence intervals at standard significance levels overlap across the simulated and actual data. Of course, in the actual data, other phenomena could lead to stochastic volatility. In the last column, we control for the prior day's realized absolute return as well as the absolute cumulative order imbalance over the first part of the day. $\lambda(t, Y_t)$ continues to predict volatility, and the magnitude of its coefficient is quite similar to that in the simulated data.

2.5 Time series of estimates

Figure 7 displays the time series of cross-sectional averages and interquartile ranges of the parameter estimates. This supplements the summary statistics given for the panel in Table 1. The average α is almost 70% in the early part of the sample and falls to about 50% by the end of the sample. This effect starts in 2007, coincident with the introduction of the NYSE Hybrid Market, which increased automated electronic execution and increased execution speeds. It is possible that market changes altered incentives to pursue private information, resulting in lower α estimates. Hendershott and Moulton (2011) find that prices became more efficient following the roll-out of the Hybrid Market, which aligns with a

Table 4
Panel regressions of end-of-day absolute returns

A. Simulated

	(1)	(2)	(3)	(4)
$\lambda(t, Y_t)$	122.90***	101.80***	50.91***	50.91***
	[17.63]	[15.14]	[9.44]	[9.44]
Constant	121.30***			
	[7.67]			
Observations	5,031,180	5,031,180	5,031,180	5,031,180
R^2	0.013	0.073	0.157	0.157
Year FEs	No	Yes	Yes	Yes
Firm FEs	No	No	Yes	Yes
Data	Simulated	Simulated	Simulated	Simulated

B. Actual

	(1)	(2)	(3)	(4)
$\lambda(t, Y_t)$	96.28***	83.81***	37.76***	48.94***
	[9.80]	[7.35]	[5.18]	[4.90]
Lag abs ret				0.15***
				[0.01]
Abs OIB				7.10***
				[0.37]
Constant	83.91***			
	[5.11]			
Observations	4,918,667	4,918,667	4,918,667	4,918,667
R^2	0.012	0.056	0.114	0.136
Year FEs	No	Yes	Yes	Yes
Firm FEs	No	No	Yes	Yes
Data	Actual	Actual	Actual	Actual

The dependent variable is the absolute return over the last 3.5 hours of the day (expressed in basis points). The model-implied price impact, $\lambda(t, Y_t)$, is defined in Equation (7) and is based on the cumulative order flow over the first 3 hours of the day. Lag Abs Ret is the absolute daily return from the previous day. Abs OIB is the absolute value of the cumulative order flow over the first 3 hours of the day. Panel A uses daily data simulated from the panel of estimated parameters for NYSE firms. Panel B uses the actual daily data. Standard errors are clustered by firm and year and are reported in brackets. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

reduced probability of private information events.[19] The other components of private information events are the magnitude κ of the signal and the likelihood p_L of a bad event. The κ estimates initially rise during the late 1990s but exhibit a strong downward trend thereafter. The average p_L indicates that the distribution of information is relatively symmetric between positive and negative events. We combine these estimates into a single composite measure of information asymmetry by calculating the expected average lambda from Equation (8). The estimates of this composite measure indicate that the amount of private

---

[19] In untabulated results, we find that the decline in α starting in 2007 is more pronounced for larger firms. Algorithmic traders (including high-frequency traders) disproportionately trade in large stocks, so it is unsurprising that the increased automation and execution speed of the Hybrid Market affected large firms more than small firms.

(a) α

(b) κ

(c) p_L

(d) σ

(e) δ
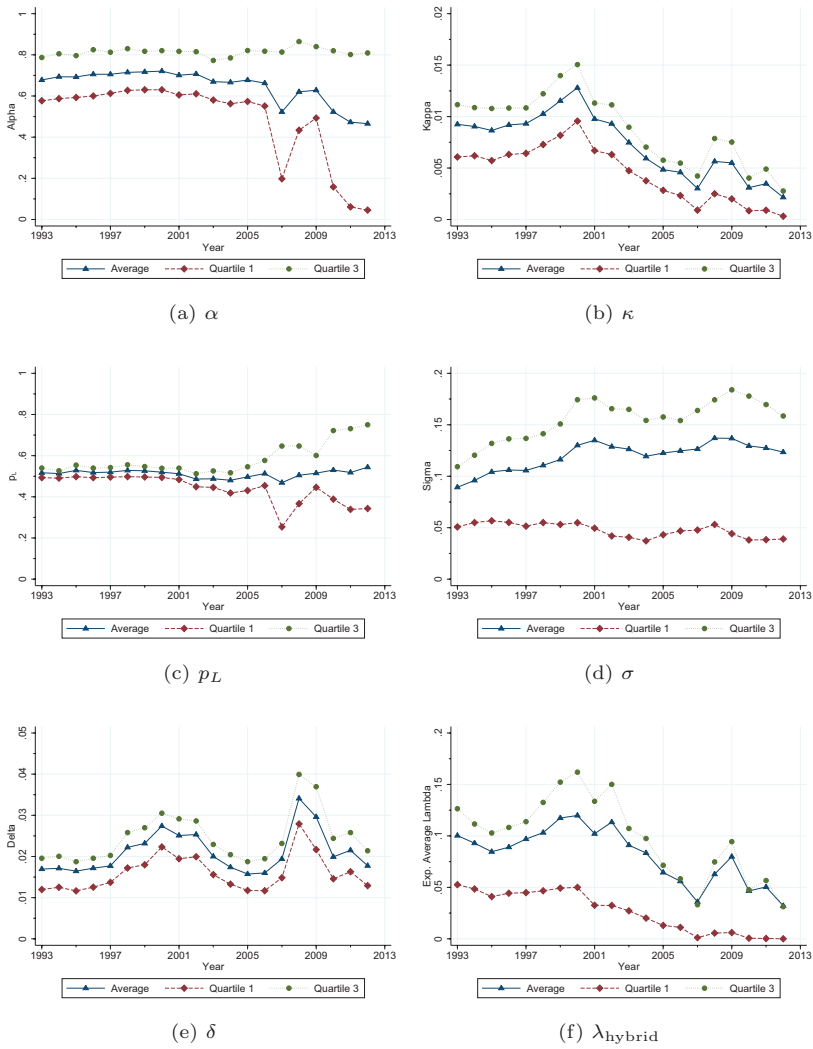
(f) $\lambda_{\mathrm{hybrid}}$

Figure 7
The annual cross-sectional mean and 25th and 75th percentiles of parameter estimates for the hybrid model
The model is estimated on a stock-year basis for NYSE stocks from 1993 to 2012 using prices and order imbalances in six hourly intraday bins and at the close. The mean and the 25th and 75th percentiles are shown. The model parameters are α = probability of an information event, κ = signal scale parameter, σ = standard deviation of liquidity trading, δ = volatility of public information, and p_L = probability of a negative event. $\lambda_{\mathrm{hybrid}}$ is the expected average lambda $\lambda(0,0)$ based on Equation (8).

information has fallen across the twenty-year sample with the exception of the late 1990s and the financial crisis.[20]

––––––––––

[20] As we will discuss in Section 4.3, the same pattern is seen in reduced-form price impact measures.

In general, the standard deviation σ of order imbalances and the volatility δ of public information appear to be roughly stationary. Despite the well-documented rise of high-frequency trading and the associated sharp increase in trading volume, the volatility of order imbalances has remained fairly stable over the twenty-year sample. Like private information, public information volatility also spiked during the financial crisis. This suggests private information may be proportional to public information rather than a fixed amount.

3. Applications

We now discuss potential applications of the estimation procedure. A large literature uses the PIN model, as discussed previously. Broadly speaking, some of this work relates PIN estimates to times when researchers believe information events have likely occurred. Other research uses PIN to proxy for information asymmetry or price informativeness. We discuss examples of how our estimates might be useful to research of either type.

3.1 Detecting information events

Information asymmetry is generally unobservable, so testing performance of adverse selection measures is challenging. In this subsection, we study how the conditional probability of an information event as measured by our model varies in two settings considered in the literature: earnings announcements and trading by Schedule 13D filers.

3.1.1 Earnings announcements. Many studies have examined the information environment surrounding earnings announcements. Some studies assume that information asymmetry is higher prior to information events, while others note that private ability or knowledge to interpret public information may result in adverse selection following announcements (Kim and Verrecchia, 1997). Several recent papers use conditional estimates based on the PIN and/or OWR models around earnings announcements (Brennan, Huh, and Subrahmanyam, 2016) and opportunistic insider trades (Duarte, Hu, and Young, 2017).

As discussed in Section 1.1, one can assess the probability of an information event if one observes cumulative order flows *and* knows the underlying parameters. In particular, Theorem 1 shows that market makers update their conditional probabilities of an information event, CPIE_t, as

$$\text{CPIE}_t(Y_t) = \begin{cases} \text{N}\left(\frac{y_L - Y_t}{\sigma\sqrt{1-t}}\right) + \text{N}\left(\frac{Y_t - y_H}{\sigma\sqrt{1-t}}\right) & \text{if } t < 1, \\ \mathbb{1}(Y_1 < y_L) + \mathbb{1}(Y_1 > y_H) & \text{if } t = 1. \end{cases} \tag{13}$$
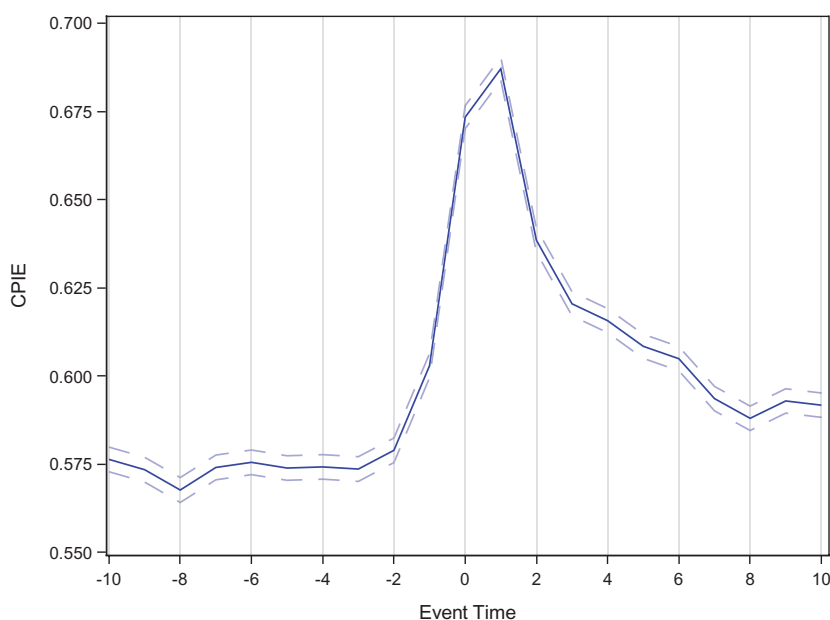
Figure 8
Averages of the end-of-day conditional probability of an information event (CPIE) in event time around earnings announcements
The CPIE is defined in Equation (13). It is calculated using the estimated parameters and order flows. Dashed lines indicate the 95% confidence interval.

Armed with our estimates of the parameters, we examine end-of-day conditional probabilities of an information event, $CPIE_1$, on the days around earnings announcements. We also calculate conditional probabilities of positive and negative information events, $CPIE^+$ and $CPIE^-$, respectively, which are the two components of CPIE in (13).

Figure 8 plots the cross-sectional average of model-implied CPIE in event time around earnings announcements. The average CPIE rises significantly on day $t-1$, consistent with early leakage of some information prior to the announcement. The average CPIE is highest on days t and $t+1$, and then falls over the next week or so. The results suggest that adverse selection may actually be worse following an earnings announcement rather than before it, as discussed in Kim and Verrecchia (1997).[21]

Pre-announcement information asymmetry is likely higher when a firm experiences an earnings surprise. To test whether CPIE captures this, we use data from IBES to calculate standardized unexpected earnings, SUE,

---

[21] This conclusion is also reached by Krinsky and Lee (1996) using the adverse selection component of bid-ask spreads and by Brennan, Huh, and Subrahmanyam (2016) using conditional probabilities from the PIN model.

calculated as

$$\text{SUE}_t = \frac{\text{EPS}_{\text{actual},t} - \text{EPS}_{\text{median forecast},t}}{P_t}, \qquad (14)$$

where $\text{EPS}_{\text{median forecast},t}$ is the median analyst forecast in the 90 days prior to the earnings announcement. We expect there to be more informed trading when the absolute value of SUE is higher. Moreover, the informed trading should correspond to the subsequent direction of the earnings surprise. That is, higher (lower) signed earnings surprises should correspond to higher CPIE$^+$ (CPIE$^-$) preceding announcements. The first three columns of Table 5 show that this is indeed the case. The average conditional probability of an information event in the 5 days preceding announcements is 80 bps higher for above median |SUE| observations relative to below median magnitude surprises. The average CPIE preceding earnings where the |SUE| is in the top decile is almost 3% higher than the average across smaller earnings surprise events. Table 5 shows that the direction of the surprises also corresponds to positive or negative event probabilities. Average CPIE$^+$ is higher before more positive SUE events, and average CPIE$^-$ is higher preceding more negative SUE events.

Greater amounts of new information also increase the likelihood that asymmetrically informed investors can trade advantageously *following* an announcement (Kim and Verrecchia, 1997). If this is the case, we expect larger magnitude |SUE| to be correlated with informed trading in the post-announcement period. Column 4 of Table 5 confirms that this is the case. In the 5 days following announcements, CPIE is higher for larger magnitude surprises. Moreover, the differences are larger than those in the pre-announcement period, again suggesting that there is more informed trading following earnings announcements than preceding them. The final two columns of Table 5 show that average CPIE$^+$ is higher following more positive surprises, while average CPIE$^-$ is higher following the most negative surprises.

3.1.2 Schedule 13D filings. Collin-Dufresne and Fos (2015) examine whether various measures of adverse selection are higher during periods in which Schedule 13D filers accumulate ownership positions. Announcement of these positions generally produces a positive stock price reaction, so these investors are privately informed. These investors must disclose days on which they traded over a 60-day period preceding the filing date. Thus, this data provides the econometrician with a laboratory concerning informed trading. Collin-Dufresne and Fos (2015) show that measures designed to capture information asymmetry are actually lower on days when Schedule 13D filers trade. As they discuss, this could be due to endogenous trading in times of greater liquidity and due to the use of patient limit orders. These effects arise in part because of the 13D filers' ability to control the timing of the private information revelation. This differs from the pre-earnings announcement setting where an informed trader's information is valid only for an exogenous duration.

Table 5
Average conditional probabilities and earnings surprises

A. Above/below median absolute or signed surprise

	Pre-announcement			Post-announcement		
	CPIE	CPIE$^+$	CPIE$^-$	CPIE	CPIE$^+$	CPIE$^-$
Top half |SUE|	0.79**			1.52***		
	(2.45)			(4.40)		
Top half SUE		0.47*			1.45***	
		(1.89)			(3.31)	
Bottom half SUE			0.60**			2.10***
			(2.24)			(6.14)

B. Top/bottom quartile absolute or signed surprise

	Pre-announcement			Post-announcement		
	CPIE	CPIE$^+$	CPIE$^-$	CPIE	CPIE$^+$	CPIE$^-$
Top quartile |SUE|	1.57***			3.06***		
	(4.48)			(8.80)		
Top quartile SUE		0.73**			2.32***	
		(2.40)			(5.87)	
Bottom quartile SUE			1.15***			3.07***
			(3.02)			(6.32)

C. Top/bottom decile absolute or signed surprise

	Pre-announcement			Post-announcement		
	CPIE	CPIE$^+$	CPIE$^-$	CPIE	CPIE$^+$	CPIE$^-$
Top decile |SUE|	2.77***			4.76***		
	(5.22)			(9.38)		
Top decile SUE		1.24***			3.29***	
		(3.63)			(6.90)	
Bottom decile SUE			1.97***			4.11***
			(3.74)			(7.51)

The conditional probability of an information event (CPIE) is defined in Equation (13). CPIE is the sum of the conditional probabilities of good and bad events, CPIE$^+$ and CPIE$^-$, respectively. The conditional probabilities are expressed as percentages. The reported estimates are the differences in average conditional probabilities of information events for the indicated quantile of absolute earnings surprises (\|SUE\|) or earnings surprise (SUE) relative to other observations. Panel A divides the sample into above and below median absolute or signed surprises. Panel B uses the top and bottom quartiles, and panel C uses the top and bottom deciles. The first three columns report the incremental averages of CPIE, CPIE$^+$, and CPIE$^-$, respectively, for the 5 days preceding the earnings announcement. The last three columns report the incremental average conditional probabilities for the 5 days following the earnings announcement. The regressions control for firm and year fixed effects, and standard errors are clustered by firm and year. *t*-statistics of the differences are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

We revisit the Schedule 13D setting to assess whether the conditional probability of an information event is higher on days when these informed investors trade. According to our model, there are informed trades on days when there are information events. So we regard the days on which 13D filers trade as information event days. Consistent with this, Collin-Dufresne and Fos (2015) show that days when Schedule 13D filers trade are characterized by significant market-adjusted returns. 13D filers typically accumulate shares by trading on occasional days over a period of weeks. Over the 60-day disclosure window, the probability that a Schedule 13D filer trades on a given day ranges from

Table 6
Average levels of the CPIE on days when Schedule 13D filers do or do not trade

	Days with informed trading	Days with no informed trading	Difference
	(1)	(2)	(3)
Full disclosure window: Days $[t-60, t-1]$			
CPIE	69.5	61.7	7.8***
			(4.86)
1st half of disclosure window: Days $[t-60, t-31]$			
CPIE	66.7	61.3	5.3**
			(2.35)
2nd half of disclosure window: Days $[t-30, t-1]$			
CPIE	71.2	62.0	9.2***
			(4.94)

The conditional probability of an information event (CPIE) is defined in Equation (13). CPIE is expressed as a percentage. The sample contains trading days in the 60-day disclosure period prior to a Schedule 13D filing date for NYSE firms in the sample of Collin-Dufresne and Fos (2015). The first column reports the average CPIE on days when Schedule 13D filers trade. The second column reports the average CPIE on days when Schedule 13D filers do not trade. The third column reports the differences between the two types of days. We report the analysis for two subperiods: the first and second halves of the disclosure period (days $[t-60, t-31]$ and $[t-30, t-1]$, respectively). Standard errors are clustered by event. t-statistics of the differences are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

around 25% to 50% (Collin-Dufresne and Fos, 2015, figure 1). One potential reason for trading on particular days is news that causes revisions in estimates of the value of activism. If activists are better informed than the market about such valuation revisions, which is quite likely, these events fit our model of private information.[22]

Table 6 reports average values of CPIE on days during the 60-day disclosure window when Schedule 13D filers do or do not trade. Just under two-thirds of the firm-days with no Schedule 13D trades are identified as being event days. On the other hand, 70% of the days when Schedule 13D filers do trade are identified as event days. The increase of 7.8% is statistically significant and represents about a 13% increase in the conditional probability relative to non-13D trading days. Thus, despite the fact that trading by Schedule 13D filers is inversely correlated with the various measures of permanent price impact commonly used in the literature and employed by Collin-Dufresne and Fos (2015), we find that the trading by 13D filers is manifested in higher conditional probabilities of an information event, calculated according to our model.

---

[22] Another reason that 13D filers may choose to trade on particular days is that liquidity trading may be time varying. This reason is proposed by Collin-Dufresne and Fos (2015). We could accommodate that by allowing σ to be time varying, but that extension is beyond the scope of the paper. Our goal here is to show that our current model, with constant σ, is informative about trading by 13D filers.

We also report average CPIE for two subperiods, the first and second halves of the disclosure period (days $[t-60, t-31]$ and $[t-30, t-1]$, respectively). If block accumulation by a 13D filer is detected by other strategic traders, then both the 13D filer and the other strategic traders should trade aggressively to beat others to the market (Holden and Subrahmanyam, 1992). This is more likely to have occurred during the second subperiod, so we expect Schedule 13D filers to trade more aggressively (use more market orders rather than limit orders) in the second subperiod. Furthermore, the second subperiod includes the period after crossing the 5% threshold, after which the 13D must be filed within ten days. We certainly expect more aggressive trading during that period. As a result of these considerations, we expect signed order flow to reflect the presence of informed trade more in the second subperiod than in the first. The second and third rows of Table 6 show that this is indeed the case. There is a smaller difference of 5.3% in CPIE over the first 30 days of the block-accumulation period between Schedule 13D trading days and nontrading days. In the second half of the disclosure period, however, the average CPIE is 9.2% higher on days when informed Schedule 13D filers trade than on days they do not.

3.2 Measuring the information content of prices

Some studies use PIN to measure the information content of prices in order to test various economic theories. Applications in corporate finance include Chen, Goldstein, and Jiang (2007), Ferreira and Laux (2007), and Bharath, Pasquariello, and Wu (2009), and applications in accounting include Frankel and Li (2004), Jayaraman (2008), and Brown and Hillegeist (2007).

Here, we demonstrate how our structural estimates could be used to augment one such study. Chen, Goldstein, and Jiang (2007) study how corporate managers learn from prices in making investment decisions. They find that investment sensitivity to prices (q) is increasing with price informativeness as proxied by PIN and by $1-R^2$ from an asset pricing model. In Table 7, we replicate Chen, Goldstein, and Jiang (2007) for our sample. Before running the regressions, we standardize each information environment variable to have unit standard deviation. Like in Chen, Goldstein, and Jiang (2007), the coefficient on q is increasing in PIN (Column 2).

To demonstrate how researchers might employ our methodology in this setting, we consider two composite measures of the information environment from the hybrid model. The first is the standard deviation of the signal (SD(ξS)) from Equation (12). We also calculate the proportion of the return variance due to private information, which we term the order-flow component of prices (OFC):

$$\frac{\text{var}(\xi S)}{\text{var}(\xi S) + \text{var}(e^{\delta B_{i1} - \delta^2/2})} = \frac{\text{SD}(\xi S)^2}{\text{SD}(\xi S)^2 + e^{\delta^2} - 1}. \tag{15}$$

Columns 4 and 5 of Table 7 show that investment-price sensitivity is increasing in each of these measures.

Table 7
Panel regressions of corporate investment

	(1)	(2)	(3)	(4)	(5)	(6)
q	1.62***	1.19***	2.08***	1.16***	1.28***	0.98***
	(8.27)	(4.67)	(7.24)	(4.50)	(5.33)	(3.11)
$q \times \text{PIN}$		0.19***				
		(2.63)				
$q \times \alpha_{\text{PIN}}$			0.00			
			(0.01)			
$q \times \frac{\varepsilon}{\mu}$			−0.29***			
			(−2.61)			
$q \times \text{SD}(\xi S)$				0.28***		
				(3.31)		
$q \times \text{OFC}$					0.22**	
					(2.44)	
$q \times \alpha_{\text{hybrid}}$						0.17***
						(3.91)
$q \times \kappa_{\text{hybrid}}$						0.26***
						(3.43)
$q \times \sigma_{\text{hybrid}}$						−0.19*
						(−1.80)
CF	7.55***	7.58***	7.72***	7.74***	7.86***	7.56***
	(5.35)	(5.37)	(5.45)	(5.49)	(5.47)	(5.43)
RET	−0.18	−0.18	−0.19	−0.16	−0.19	−0.19
	(−1.52)	(−1.49)	(−1.62)	(−1.48)	(−1.64)	(−1.64)
INV ASSET	0.56***	0.52**	0.51**	0.55***	0.52**	0.46**
	(2.72)	(2.57)	(2.51)	(2.67)	(2.53)	(2.29)
PIN		−0.23***				
		(−2.73)				
α_{PIN}			0.01			
			(0.11)			
$\frac{\varepsilon}{\mu}$			0.31**			
			(2.20)			
SD(ξS)				−0.52***		
				(−4.04)		
OFC					−0.16	
					(−1.38)	
α_{hybrid}						−0.22***
						(−3.36)
κ_{hybrid}						−0.40***
						(−3.68)
σ_{hybrid}						−0.32
						(−1.41)
Adjusted R^2	0.745	0.746	0.746	0.747	0.746	0.748
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes

The dependent variable is capital expenditures. The independent variable q is market-to-book of assets. PIN is the probability of informed trading from Easley et al. (1996). SD(ξS) is the standard deviation of the signal ξS like in Equation (12). OFC is the proportion of return variance due to private information (the order-flow component of prices) like in Equation (15). α is the probability of an information event in either the PIN or the hybrid model. κ_{hybrid} is the magnitude of an information event and σ_{hybrid} is the standard deviation of liquidity trading from the hybrid model. ε/μ is the ratio of the liquidity to informed trading intensities from PIN. Each information environment variable is standardized to have unit standard deviation. CF is firm cash flows. RET is the cumulative return over the next three years. INV ASSET is the inverse of the book value of assets. Standard errors are clustered by firm and year. t-statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

One advantage of our estimation procedure relative to PIN is that it allows us to separately estimate the probability and magnitude of information events. Investment sensitivity to prices is increasing in each of these components (Column 6 of Table 7). Thus, when there are more frequent or larger episodes of private information, investment is more sensitive to prices. A 1-standard-deviation increase in κ (the magnitude of information events) is associated with about a 25% increase in investment-price sensitivity. A standard deviation change in α (the probability of an information event) has an effect about two-thirds as large. The positive effect of α conflicts with results from decomposing PIN into the probability of an information event and the relative intensity of liquidity to informed traders (Column 3). An increase in the PIN α does not lead to increased investment sensitivity to prices.

3.3 Probability and magnitudes of private information

Estimation of the probability and magnitude of information events could also prove useful in other settings where researchers are interested in the information environment. For instance, the estimates can provide additional texture to studies of the effects of information-related regulation such as insider trading laws, short-selling restrictions, or symmetric access to managers for financial analysts (e.g., Reg FD in the United States). Separating the probability and magnitude of information events could be useful in the analyst literature more broadly. Do analysts turn private information into public information? If so, one might expect to see lower probabilities of information events for firms with greater analyst coverage. On the other hand, analysts may produce private information, which could result in higher probabilities of information events. Studies interested in how the investor base affects liquidity could be more nuanced by including both α and κ. Index inclusion affects institutional ownership, so how does index inclusion affect the information environment? Greater institutional ownership could result in lower magnitudes of private information if prices are more efficient with institutional ownership. The accounting literature considers whether disclosure quality and frequency affect the information environment of firms. Greater disclosure quality could reduce the magnitude of private information, and greater disclosure frequency could reduce the probability of private information events. In all of these cases, studying both α and κ could improve our understanding relative to studying only composite measures of private information.

4. Comparison to Other Models

In this section, we compare the estimates of our model to those of the three structural models (PIN, APIN, OWR) and the reduced-form version of PIN (VPIN) discussed in the Introduction. The estimation procedure for the other models is detailed in Internet Appendix C.

Table 8
Correlations of structural parameters from the hybrid and other models

A. Composite measures

	λ_{hybrid}	PIN	λ_{OWR}	APIN	VPIN
λ_{hybrid}	1.00				
PIN	0.35	1.00			
λ_{OWR}	0.55	0.17	1.00		
APIN	0.42	0.58	0.19	1.00	
VPIN	0.56	0.42	0.26	0.48	1.00

B. Probability of an information event

	α_{hybrid}	α_{PIN}	α_{OWR}	α_{APIN}	VPIN
α_{hybrid}	1.00				
α_{PIN}	−0.09	1.00			N/A
α_{OWR}	−0.09	0.05	1.00		
α_{APIN}	−0.01	0.25	0.04	1.00	

C. Liquidity trading

	σ_{hybrid}	$\frac{\varepsilon}{\mu}$	σ_u	$\frac{\varepsilon+\theta\eta}{\mu}$	VPIN
σ_{hybrid}	1.00				
$\frac{\varepsilon}{\mu}$	0.57	1.00			N/A
σ_u	0.92	0.51	1.00		
$\frac{\varepsilon+\theta\eta}{\mu}$	0.53	0.83	0.48	1.00	

For all models, α = probability of an information event. For the hybrid model, λ_{hybrid} is the expected average lambda $\lambda(0,0)$ based on Equation (8). PIN, APIN, and VPIN are the probabilities of informed trading estimated using the methodologies in Easley et al. (1996), Duarte and Young (2009), and Easley, López de Prado, and O'Hara (2012), respectively. λ_{OWR} is the estimate of Kyle's lambda from Odders-White and Ready (2008). σ_{hybrid} and σ_u are the standard deviations of liquidity trading from the hybrid and OWR models, respectively. ε/μ and $(\varepsilon+\theta\eta)/\mu$ are the ratios of the liquidity to informed trading intensities from the PIN and APIN models, respectively.

4.1 Correlations of model parameters

Panel A of Table 8 shows the correlations among PIN, APIN, VPIN, lambda from the OWR model (λ_{OWR}), and the expected average lambda from our model (λ_{hybrid}) – see Equation (8). All of the correlations are positive. The largest correlations with λ_{hybrid} are those of the OWR lambda and VPIN. This is perhaps not surprising since each of these estimates uses price changes in some form. The OWR lambda uses the joint distribution of returns and order flows, while VPIN signs volume using price changes.

We call PIN, APIN, VPIN, λ_{owr}, and λ_{hybrid} composite measures of information asymmetry because, with the exception of VPIN, they are functions of the underlying structural parameters.[23] We also examine the correlations of the structural parameters of the various models. Panel B of Table 8 reports correlations of the estimated probability of an information event from each model (except VPIN which does not identify α). The estimates of α for the hybrid model are negatively correlated with estimates of α from the other

---

[23] We refer to VPIN as reduced form because it does not identify the underlying structural parameters. Rather, it proxies for PIN by separately estimating the numerator and denominator of PIN (see Internet Appendix C.4).

models. In each of the other models, the unconditional distribution of order flow imbalances changes with α, unlike in our model, so the lack of correlation of the hybrid model α with the other α's is consistent with the identification discussion in Section 1.1. The implications of the models for the unconditional distribution of order flow imbalances are discussed further in Internet Appendix D.[24]

The positive correlation of λ_{hybrid} with the other composite measures is somewhat surprising given that the α of the hybrid model is not positively correlated with the α's of the other models. The explanation lies in the estimates of liquidity trading. Equation (8) shows that the expected average lambda is inversely related to the volatility of liquidity trading. The other measures are also inversely related to liquidity trading (see Equations (C.2), (C.4), and (C.6) in the Internet Appendix). Panel C of Table 8 reports correlations of the liquidity trading parameters of each model. We scale the PIN and APIN liquidity trading parameters by the estimated μ, so the fractions ε/μ and $(\varepsilon+\theta\eta)/\mu$ represent the intensity of liquidity trading relative to informed trading. Note that PIN and APIN are decreasing in these ratios, respectively. The liquidity trading parameters are positively correlated across the models. For this reason, the composite measures are positively correlated despite the lack of correlation of the estimated alphas.

4.2 Cross-sectional variation in parameters

It is interesting to see how estimates of private information differ in the cross-section of firms across models. Table 9 reports average values of the estimates within market capitalization deciles. Across all of the models, composite measures of information asymmetry decrease in firm size (panel A). For the hybrid model, the average probability α of an information event decreases in firm size, whereas the estimates for the other models are exactly the opposite, increasing in firm size (panel B). Like in the unconditional correlation analysis, the composite measures seem to behave similarly in the size cross-section due to similarities in liquidity trading measurement (panel C). Estimates from all of the models indicate more intense liquidity trading for larger capitalization stocks. For each of the models other than the hybrid model, the effect of the more pronounced liquidity trading dominates the modest increases in α as a function of size, so these composite measures are lower for larger firms as a result of higher estimated liquidity trading.[25]

4.3 Relation to price impacts and quoted spreads

In theory, price impacts and quoted spreads should be larger when information asymmetry is higher. This is shown in Section 1 for price impacts in the hybrid

---

[24] Venter and de Jongh (2006), Duarte and Young (2009), Gan, Wei, and Johnstone (2014), and Duarte, Hu, and Young (2017) all show that the PIN model fails to fit the empirical joint distribution of buy and sell orders.

[25] The OWR lambda is also a function of its estimated magnitude of private information σ_i. For both the hybrid model and the OWR model, the estimated magnitude of private information is also decreasing in size.
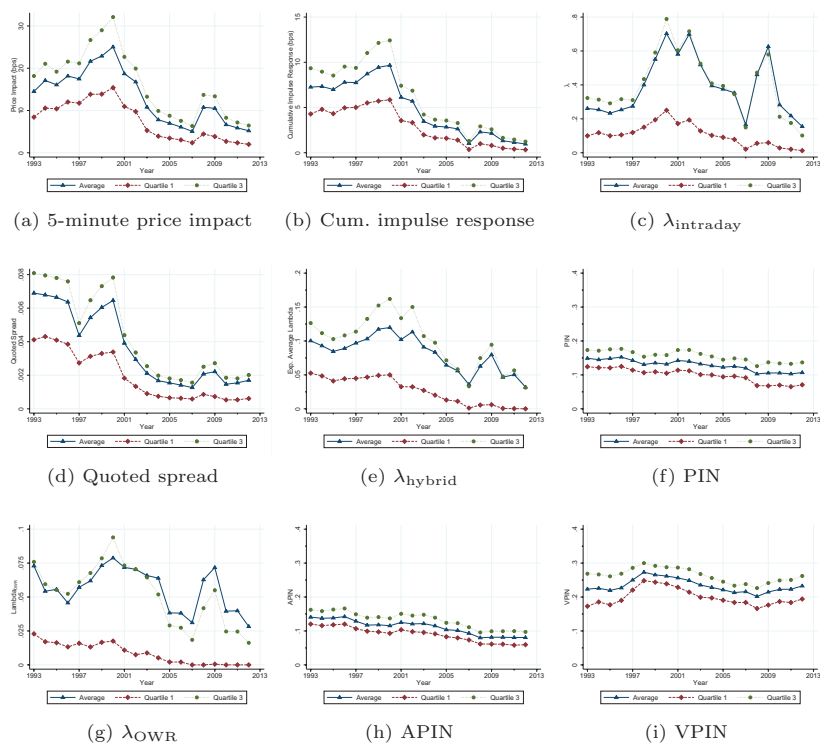
Table 9
Average values of parameter estimates within market capitalization deciles

A. Composite measures

	λ_{hybrid}	PIN	λ_{OWR}	APIN	VPIN
1 (Small)	0.200	0.18	0.139	0.15	0.28
2	0.144	0.15	0.089	0.13	0.27
3	0.111	0.14	0.068	0.12	0.25
4	0.085	0.13	0.058	0.12	0.24
5	0.066	0.13	0.048	0.11	0.23
6	0.052	0.12	0.040	0.10	0.23
7	0.042	0.12	0.034	0.10	0.22
8	0.035	0.11	0.032	0.09	0.21
9	0.025	0.09	0.024	0.08	0.20
10 (Large)	0.020	0.08	0.020	0.07	0.18

B. Probability of an information event

	α_{hybrid}	α_{PIN}	α_{OWR}	α_{APIN}	VPIN
1 (Small)	0.74	0.31	0.11	0.41	
2	0.71	0.33	0.12	0.44	
3	0.69	0.34	0.12	0.44	
4	0.67	0.35	0.12	0.45	
5	0.65	0.36	0.14	0.45	N/A
6	0.63	0.36	0.14	0.45	
7	0.62	0.38	0.15	0.46	
8	0.59	0.38	0.17	0.46	
9	0.56	0.39	0.18	0.46	
10 (Large)	0.52	0.39	0.23	0.47	

C. Liquidity trading

	σ_{hybrid}	$\frac{\varepsilon}{\mu}$	σ_u	$\frac{\varepsilon+\theta\eta}{\mu}$	VPIN
1 (Small)	0.06	0.73	0.04	1.24	
2	0.06	0.94	0.04	1.51	
3	0.07	1.06	0.05	1.69	
4	0.08	1.19	0.06	1.84	
5	0.09	1.28	0.08	1.97	N/A
6	0.11	1.38	0.09	2.08	
7	0.12	1.55	0.11	2.26	
8	0.15	1.74	0.14	2.50	
9	0.19	2.13	0.19	2.83	
10 (Large)	0.29	2.64	0.33	3.42	

Stocks are sorted into capitalization deciles annually. For all models, α = probability of an information event. For the hybrid model, λ_{hybrid} is the expected average lambda $\lambda(0,0)$ based on Equation (8). PIN, APIN, and VPIN are the probabilities of informed trading estimated using the methodologies in Easley et al. (1996), Duarte and Young (2009), and Easley, López de Prado, and O'Hara (2012), respectively. λ_{OWR} is the estimate of Kyle's lambda from Odders-White and Ready (2008). σ_{hybrid} and σ_u are the standard deviations of liquidity trading from the hybrid and OWR models, respectively. ε/μ and $(\varepsilon+\theta\eta)/\mu$ are the ratios of the liquidity to informed trading intensities from the PIN and APIN models, respectively.

model. For the PIN model, the opening quoted spread is the product of PIN and the magnitude of the information, $H - L$.[26] In this section, we assess how time-series and cross-sectional variations in price impacts and quoted spreads relate to the estimated composite measure from each model. For price impacts, we use the three measures described in Section 2.3. Quoted spreads are the time-weighted average proportional bid-ask spreads.

---

[26] See Equation (11) of Easley et al. (1996), who assume $p_L = p_H$.

(a) 5-minute price impact   (b) Cum. impulse response   (c) $\lambda_{\text{intraday}}$

(d) Quoted spread   (e) λ_{hybrid} (f) PIN

(g) λ_{OWR} (h) APIN   (i) VPIN

Figure 9
The annual cross-sectional mean and the 25th and 75th percentiles of reduced-form price impacts, quoted spreads, and composite information asymmetry measures
Five-minute price impacts are estimated daily and averaged annually for each stock-year for NYSE stocks from 1993 to 2012. The stock-year estimates of the cumulative impulse response and $\lambda_{\text{intraday}}$ are the medians of daily estimates. Quoted spread is the time-weighted proportional bid-ask spread. λ_{hybrid} is the expected average lambda $\lambda(0,0)$ based on Equation (8). PIN, APIN, and VPIN are the probabilities of informed trading estimated using the methodologies in Easley et al. (1996), Duarte and Young (2009), and Easley, López de Prado, and O'Hara (2012), respectively. λ_{OWR} is the estimate of Kyle's lambda from Odders-White and Ready (2008).

Figure 9 plots the time series of the cross-sectional averages and interquartile ranges of the price impact measures, the quoted spread, and the five composite information asymmetry measures. Over the twenty year sample, price impacts initially rose over the 1990s before falling dramatically following the turn of the century, with the brief exception of the financial crisis. Quoted spreads have also fallen over the sample period. The time series of the hybrid model expected average lambda, λ_{hybrid}, and the magnitude of private information, κ, exhibit similar patterns (Figure 7). The OWR lambda also exhibits similar behavior. PIN, APIN, and VPIN are much less variable over time.

Table 10 explores the time-series relationships across these measures more formally. For each firm with at least five years of estimates, we calculate the time-series correlations between the price impact or quoted spread measure

Table 10
Time-series correlations of reduced-form and structural estimates

A. Average time-series correlations

	5-minute price impact	Cum. impulse response	$\widehat{\lambda}_{\text{intraday}}$	Quoted spread
λ_{hybrid}	0.641	0.702	0.584	0.619
PIN	0.297	0.327	0.238	0.346
λ_{OWR}	0.331	0.343	0.309	0.331
APIN	0.379	0.448	0.310	0.449
VPIN	0.513	0.520	0.407	0.441

B. t-statistics of paired t-tests of differences

	5-minute price impact	Cum. impulse response	$\widehat{\lambda}_{\text{intraday}}$	Quoted spread
PIN	30.5***	34.1***	30.7***	23.5***
λ_{OWR}	33.5***	39.9***	29.3***	30.6***
APIN	24.3***	23.8***	25.6***	15.1***
VPIN	11.0***	15.7***	15.7***	13.7***

The table reports cross-sectional averages of the time-series correlation between reduced-form liquidity estimates (each column) and the composite structural information asymmetry variables (each row). The reduced-form liquidity variables are the 5-minute price impact, the cumulative impulse response estimated following Hasbrouck (1991), an estimate of price impact ($\widehat{\lambda}_{\text{intraday}}$) using a regression of 5-minute returns on the square root of signed volume following Hasbrouck (2009) and Goyenko, Holden, and Trzcinka (2009), and the proportional quoted spread. The time-series correlation is calculated for each firm with at least five years of observations. Panel A reports the cross-sectional average of the time-series correlations. Panel B reports *t*-statistics of paired *t*-tests of the time-series correlation of λ_{hybrid} with the variable in the column header relative to the corresponding correlation for the composite variable in each row.

and each model-based composite measure. Table 10 reports the cross-sectional average of these time-series correlations. For all three reduced-form price impact estimates and for quoted spreads, λ_{hybrid} is the most correlated composite measure and is significantly more correlated than the other composite measures. Using the approximately 1,600 firms with at least five years of estimates, paired *t*-tests reject the nulls that the correlation with λ_{hybrid} equals the correlations with the other composite measures (panel B of Table 10).

We also explore how the composite measures relate cross-sectionally to the price impact and quoted spread benchmarks. Table 11 reports cross-sectional regressions of price impacts and quoted spreads on the composite information asymmetry measures. We run univariate regressions as well as bivariate regressions including λ_{hybrid} and another composite measure. The information asymmetry measures are standardized to have unit standard deviations. In univariate regressions, the reduced-form price impact measures and quoted spreads are positively related to each of the information asymmetry measures. λ_{hybrid} generally explains the most (or second-most) cross-sectional variation in price impacts and explains over a quarter of the variation in quoted spreads.[27] Perhaps more importantly, λ_{hybrid} adds explanatory power to each of the other

[27] For the univariate quoted spreads regressions, VPIN has the largest average R^2, but its coefficient estimate is insignificant. This is because VPIN and quoted spreads are negatively correlated cross-sectionally over the first 5 years of the sample.

Table 11
Fama and MacBeth (1973) cross-sectional regressions of price impacts and quoted spreads

A. 5-minute price impact

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
λ_{hybrid}	0.47***		0.39***		0.48***		0.37***		0.30**
	(9.10)		(7.48)		(8.58)		(5.68)		(2.52)
PIN		0.37***	0.25***						
		(9.37)	(8.10)						
λ_{OWR}				0.26***	−0.01				
				(5.25)	(−0.70)				
APIN						0.43***	0.30***		
						(8.17)	(7.85)		
VPIN								0.41***	0.28**
								(3.95)	(2.30)
Constant	0.05	0.06	0.04	0.08	0.05	0.09	0.06	0.08	0.07
	(0.25)	(0.28)	(0.20)	(0.31)	(0.24)	(0.46)	(0.33)	(0.40)	(0.36)
Obs	19,965	19,965	19,965	19,965	19,965	19,965	19,965	19,965	19,965
R^2	0.317	0.200	0.400	0.097	0.320	0.255	0.421	0.356	0.474

B. Cumulative impulse response

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
λ_{hybrid}	0.48***		0.42***		0.50***		0.41***		0.36**
	(4.34)		(4.02)		(4.23)		(3.63)		(2.57)
PIN		0.32***	0.20***						
		(5.28)	(4.81)						
λ_{OWR}				0.26***	−0.03**				
				(3.32)	(−2.19)				
APIN						0.36***	0.23***		
						(7.22)	(7.27)		
VPIN								0.38***	0.23***
								(5.71)	(4.28)
Constant	0.07	0.08	0.05	0.12	0.07	0.07	0.04	0.07	0.05
	(0.23)	(0.27)	(0.17)	(0.35)	(0.22)	(0.27)	(0.15)	(0.27)	(0.20)
Obs	19,965	19,965	19,965	19,965	19,965	19,965	19,965	19,965	19,965
R^2	0.419	0.205	0.490	0.120	0.423	0.263	0.507	0.396	0.548

(continued)

composite measures regardless of the benchmark when comparing the bivariate and univariate regressions. This is true for both the price impact benchmarks and for quoted spreads.

The hybrid model parameters are estimated using a sample of prices and order flows, so it is perhaps unsurprising that λ_{hybrid} captures reduced-form price impacts well. However, this critique does not apply to quoted spreads, which are not part of the data used in the estimation. Tables 10 and 11 show that λ_{hybrid} also performs well vis-a-vis alternative composite measures when quoted spreads are used as the benchmark.

Of course, there remains unexplained variation in both reduced-form price impacts and quoted spreads. Some empirical work on information asymmetry has aggregated various empirical proxies of information asymmetry to try to capture the multifaceted nature of liquidity (e.g., Bharath, Pasquariello, and Wu, 2009, Korajczyk and Sadka, 2008). That none of the composite measures, including λ_{hybrid}, completely explains price impacts or quoted spreads, lends

Table 11
Continued

C. $\hat{\lambda}_{intraday}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
λ_{hybrid}	0.35***		0.27***		0.35***		0.23***		0.15
	(12.90)		(10.27)		(13.53)		(6.22)		(1.27)
PIN		0.31***	0.23***						
		(5.50)	(4.72)						
λ_{OWR}				0.20***	0.00				
				(7.16)	(0.37)				
APIN						0.41***	0.32***		
						(4.08)	(3.51)		
VPIN								0.35**	0.30
								(2.23)	(1.38)
Constant	−0.03	−0.00	−0.02	−0.02	−0.03	0.05	0.02	0.06	0.06
	(−0.47)	(−0.03)	(−0.31)	(−0.23)	(−0.45)	(0.50)	(0.24)	(0.75)	(0.67)
Obs	19,965	19,965	19,965	19,965	19,965	19,965	19,965	19,965	19,965
R^2	0.191	0.115	0.245	0.066	0.194	0.153	0.270	0.185	0.332

D. Quoted spread

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
λ_{hybrid}	0.42***		0.34***		0.42***		0.31***		0.34**
	(6.51)		(5.80)		(6.71)		(4.58)		(2.24)
PIN		0.37***	0.27***						
		(6.97)	(5.82)						
λ_{OWR}				0.24***	−0.00				
				(4.32)	(−0.29)				
APIN						0.44***	0.34***		
						(11.32)	(10.42)		
VPIN								0.19	0.06
								(1.27)	(0.34)
Constant	0.10	0.09	0.07	0.13	0.10	0.08	0.06	0.15	0.13
	(0.39)	(0.36)	(0.31)	(0.43)	(0.38)	(0.40)	(0.31)	(0.56)	(0.53)
Obs	19,965	19,965	19,965	19,965	19,965	19,965	19,965	19,965	19,965
R^2	0.257	0.204	0.353	0.081	0.259	0.279	0.390	0.347	0.461

The dependent variables in panels A–D are the 5-minute price impact, the cumulative impulse response estimated following Hasbrouck (1991), an estimate of price impact ($\hat{\lambda}_{intraday}$) using a regression of 5-minute returns on the square root of signed volume following Hasbrouck (2009) and Goyenko, Holden, and Trzcinka (2009), and the proportional quoted spread, respectively. Each panel reports univariate and bivariate regressions. All variables are standardized to have a unit standard deviation. The reported R^2 is the time-series average R^2 from the cross-sectional regressions. Standard errors are adjusted for serial correlation following Newey and West (1987) with 5 lags. *t*-statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

credence to such aggregations. Our results suggest that λ_{hybrid} or its underlying structural parameters should be included when empirical researchers wish to aggregate information asymmetry estimates.

5. Conclusion

We propose a model of informed trading that is a hybrid of the PIN and Kyle models. Unlike the Kyle model, information events occur with probability less than one like in the PIN model, and unlike the PIN model, informed orders are endogenously determined like in the Kyle model. An important implication of the model is that both returns and order flows are needed to identify

information asymmetry parameters. The reason is that order flows depend on market liquidity, which depends on information asymmetry. This is an indirect dependence of order flows on information asymmetry that is countervailing to the direct relation. This result suggests that measures of information asymmetry based solely on order flows (like PIN) may be misspecified.

We estimate the hybrid model and provide several analyses that suggest the estimates capture cross-sectional and time-series variation in information asymmetry. We illustrate possible applications of our estimates: a new methodology to detect information events and a corporate finance application. Our model allows the econometrician to identify distinct components of information asymmetry such as the probability and magnitude of potential information events. We hope such refinements will prove useful to future finance and accounting research.

Finally, we compare the parameter estimates to those from other structural models and to price impacts and quoted spreads. While composite information asymmetry measures from all of the models are positively correlated with price impacts, the measure from the hybrid model exhibits higher time-series correlations and incremental cross-sectional explanatory power for price impacts. To a certain extent, this might be expected, since the measure from the hybrid model is the expected average Kyle's lambda, and Kyle's lambda should be highly correlated with price impacts. However, the measure from the Odders-White and Ready (2008) model is also an estimate of a Kyle's lambda, and it is dominated by the hybrid model in explaining both time-series and cross-sectional variation in price impacts. Moreover, the hybrid model measure is also more correlated with quoted spreads than other measures in the time series and adds explanatory power to each of the other measures in explaining the cross-section of quoted spreads.

Appendix A. Proofs

The process Y described in the following lemma is a variation of a Brownian bridge. It differs from a Brownian bridge in that the endpoint is not uniquely determined but instead is determined only to lie in an interval: the lower tail $(-\infty, y_L)$, the upper tail (y_H, ∞), or the middle region $[y_L, y_H]$, depending on whether there is an information event and whether the news is good or bad. Part (C) of the lemma immediately follows from the preceding parts, because the probability (A3) is the probability that $Y_1 \notin [y_L, y_H]$ calculated on the basis that Y is an $\mathbb{F}^Y$-Brownian motion with zero drift and standard deviation σ.

Lemma. Let N denote the standard normal distribution function. Let $\mathbb{F}^Y = \{\mathcal{F}_t^Y \mid 0 \leq t \leq 1\}$ denote the filtration generated by the stochastic process Y defined by $Y_0 = 0$ and

$$dY_t = \frac{q(t, Y_t, \xi S)}{1-t} dt + dZ_t. \tag{A1}$$

Then the following are true:

(A) Y is an $\mathbb{F}^Y$–Brownian motion with zero drift and standard deviation σ.

(B) With probability one,

$$\xi = 1 \text{ and } S = L \quad \Rightarrow \quad Y_1 < y_L, \tag{A2a}$$

$$\xi = 0 \quad \Rightarrow \quad y_L \le Y_1 \le y_H, \tag{A2b}$$

$$\xi = 1 \text{ and } S = H \quad \Rightarrow \quad Y_1 > y_H. \tag{A2c}$$

(C) For each $t < 1$, the probability that $\xi = 1$ conditional on $\mathcal{F}_t^Y$ is

$$\mathrm{N}\left(\frac{y_L - Y_t}{\sigma\sqrt{1-t}}\right) + 1 - \mathrm{N}\left(\frac{y_H - Y_t}{\sigma\sqrt{1-t}}\right). \tag{A3}$$

■

Proof of Lemma

Set

$$k(1, y, s) = \begin{cases} 1_{\{y < y_L\}} & \text{if } s = L, \\ 1_{\{y_L \le y \le y_H\}} & \text{if } s = 0, \\ 1_{\{y > y_H\}} & \text{if } s = H, \end{cases}$$

and, for $t < 1$,

$$k(t, y, s) = \begin{cases} \mathrm{N}\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) & \text{if } s = L, \\ \mathrm{N}\left(\frac{y_H - y}{\sigma\sqrt{1-t}}\right) - \mathrm{N}\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) & \text{if } s = 0, \\ \mathrm{N}\left(\frac{y - y_H}{\sigma\sqrt{1-t}}\right) & \text{if } s = H. \end{cases}$$

Define

$$\ell(t, y, s) = \frac{\partial \log k(t, y, s)}{\partial y},$$

for $t < 1$. Then $(1-t)\sigma^2 \ell(t, y, s) = q(t, y, s)$ for $t < 1$, and the stochastic differential equation (A1) can be written as

$$\mathrm{d}Y_t = \sigma^2 \ell(t, Y_t, \xi S)\mathrm{d}t + \mathrm{d}Z_t \tag{A4}$$

The process Y is an example of a Doob h-transform—see Rogers and Williams (2000).

To put (A4) in a more standard form, define the two-dimensional process $\hat{Y}_t = (\xi S, Y_t)$ with random initial condition $\hat{Y}_0 = (\xi S, 0)$, and augment (A4) with the equation $\mathrm{d}(\xi S) = 0$. The existence of a unique strong solution $\hat{Y}$ to this enlarged system follows from Lipschitz and growth conditions satisfied by ℓ. See Karatzas and Shreve (1988, theorem 5.2.9).

The uniqueness in distribution of weak solutions of stochastic differential equations (Karatzas and Shreve, 1988, theorem 5.3.10) implies that we can demonstrate Properties (A) and (B) by exhibiting a weak solution for which they hold. To construct such a weak solution, define a new measure $\mathbb{Q}$ on $\mathcal{F}_1$ using $k(1, Z_1, \xi S)/k(0, 0, \xi S)$ as the Radon–Nikodym derivative. The definition of k implies that $k(t, Z_t, \xi S)$ is the $\mathcal{F}_t$–conditional expectation of the indicator function $k(1, Z_1, \xi S)$, so $k(t, Z_t, \xi S)$ is a martingale on the filtration $\mathbb{F}$. By Girsanov's theorem, the process Z^* defined by $Z_0^* = 0$ and

$$\mathrm{d}Z_t^* = -\sigma^2 \ell(t, Z_t, \xi S)\mathrm{d}t + \mathrm{d}Z_t$$

is a Brownian motion (with zero drift and standard deviation σ) on the filtration $\mathbb{F}$ relative to $\mathbb{Q}$. It follows that Z is a weak solution of (A4) relative to the Brownian motion Z^* on the filtered probability space $(\Omega, \mathbb{F}, \mathbb{Q})$.

To establish property (A) for the weak solution, we need to show that Z is a Brownian motion on $(\Omega, \mathbb{G}, \mathbb{Q})$. Because Z is a Brownian motion on $(\Omega, \mathbb{G}, \mathbb{P})$, it suffices to show that $\mathbb{Q} = \mathbb{P}$ when both are restricted to $\mathcal{G}_1$. This holds if for all $t_1 < \cdots < t_n \leq 1$ and all Borel B we have

$$\mathbb{P}((Z_{t_1}, \ldots, Z_{t_n}) \in B) = \mathbb{Q}((Z_{t_1}, \ldots, Z_{t_n}) \in B). \tag{A5}$$

The right-hand side of (A5) equals

$$\mathsf{E}\left[\frac{k(1, Z_1, \xi S)}{k(0, 0, \xi S)} \mathbf{1}_B(Z_{t_1}, \ldots, Z_{t_n})\right],$$

which can be represented as the following sum:

$$\alpha p_L \mathsf{E}\left[\frac{k(1, Z_1, \xi S)}{k(0, 0, \xi S)} \mathbf{1}_B(Z_{t_1}, \ldots, Z_{t_n}) \,|\, \xi S = L\right]$$

$$+ (1 - \alpha) \mathsf{E}\left[\frac{k(1, Z_1, \xi S)}{k(0, 0, \xi S)} \mathbf{1}_B(Z_{t_1}, \ldots, Z_{t_n}) \,|\, \xi = 0\right]$$

$$+ \alpha p_H \mathsf{E}\left[\frac{k(1, Z_1, \xi S)}{k(0, 0, \xi S)} \mathbf{1}_B(Z_{t_1}, \ldots, Z_{t_n}) \,|\, \xi S = H\right].$$

Using the definitions of y_L, y_H, and k, this equals

$$\mathsf{E}\left[\mathbf{1}_{\{Z_1 < y_L\}} \mathbf{1}_B(Z_{t_1}, \ldots, Z_{t_n}) \,|\, \xi S = L\right]$$

$$+ \mathsf{E}\left[\mathbf{1}_{\{y_L \leq Z_1 \leq y_L\}} \mathbf{1}_B(Z_{t_1}, \ldots, Z_{t_n}) \,|\, \xi = 0\right]$$

$$+ \mathsf{E}\left[\mathbf{1}_{\{Z_1 > y_H\}} \mathbf{1}_B(Z_{t_1}, \ldots, Z_{t_n}) \,|\, \xi S = H\right].$$

The $\mathbb{P}$–independence of Z and ξS imply that the conditional expectations equal the unconditional expectations, so adding the three terms gives

$$\mathsf{E}\left[\mathbf{1}_B(Z_{t_1}, \ldots, Z_{t_n})\right] = \mathbb{P}((Z_{t_1}, \ldots, Z_{t_n}) \in B).$$

This completes the proof that Z is a Brownian motion on $(\Omega, \mathbb{G}, \mathbb{Q})$.

To establish property (B) for the weak solution of (A4), we need to show that

$$\mathbb{Q}(Z_1 < y_L \,|\, \xi S = L) = 1, \tag{A6a}$$

$$\mathbb{Q}(y_L \leq Z_1 \leq y_H \,|\, \xi = 0) = 1, \tag{A6b}$$

$$\mathbb{Q}(Z_1 > y_H \,|\, \xi S = H) = 1. \tag{A6c}$$

Consider (A6a). We have

$$\mathbb{Q}(\xi S = L) = \mathsf{E}\left[\frac{k(1, Z_1, \xi S)}{k(0, 0, \xi S)} \mathbf{1}_{\{\xi S = L\}}\right]$$

$$= \mathsf{E}\left[\frac{k(1, Z_1, L)}{k(0, 0, L)} \mathbf{1}_{\{\xi S = L\}}\right]$$

$$= \mathsf{E}\left[\mathbf{1}_{\{Z_1 < y_L\}} \mathbf{1}_{\{\xi S = L\}}\right] \big/ \alpha p_L$$

$$= \alpha p_L,$$

using the definition of k for the third equality and the $\mathbb{P}$–independence of Z and ξS for the last equality. By similar reasoning,

$$
\mathbb{Q}(Z_1 < y_L, \xi S = L) = \mathsf{E}\left[\frac{k(1, Z_1, \xi S)}{k(0, 0, \xi S)} \mathbf{1}_{\{Z_1 < y_L\}} \mathbf{1}_{\{\xi S = L\}}\right]
$$

$$
= \mathsf{E}\left[\frac{k(1, Z_1, L)}{k(0, 0, L)} \mathbf{1}_{\{Z_1 < y_L\}} \mathbf{1}_{\{\xi S = L\}}\right]
$$

$$
= \mathsf{E}\left[\mathbf{1}_{\{Z_1 < y_L\}} \mathbf{1}_{\{\xi S = L\}}\right] / \alpha p_L
$$

$$
= \alpha p_L .
$$

Thus,

$$
\mathbb{Q}(Z_1 < y_L \,|\, \xi S = L) = \frac{\mathbb{Q}(Z_1 < y_L, \xi S = L)}{\mathbb{Q}(\xi S = L)} = \frac{\alpha p_L}{\alpha p_L} = 1.
$$

Conditions (A6b) and (A6c) can be verified by the same logic. ∎

Proof of Theorem 1

It is explained in the text why the equilibrium condition (1) holds. It remains to show that the strategy (5) is optimal for the informed trader. Let $\mathbb{G} \stackrel{\text{def}}{=} \{\mathcal{G}_t \,|\, 0 \le t \le T\}$ denote the completion of the filtration generated by Z, form the enlarged filtration with σ–fields $\mathcal{G}_t \vee \sigma(\xi S)$, and let $\mathbb{F} \stackrel{\text{def}}{=} \{\mathcal{F}_t \,|\, 0 \le t \le T\}$ denote the completion of the enlarged filtration. The filtration $\mathbb{F}$ represents the informed trader's information.

Define

$$
J(1, y, L) = -L(y - y_L)\mathbf{1}_{\{y > y_L\}} + H(y - y_H)\mathbf{1}_{\{y > y_H\}},
$$

$$
J(1, y, 0) = -L(y_L - y)\mathbf{1}_{\{y < y_L\}} + H(y - y_H)\mathbf{1}_{\{y > y_H\}},
$$

$$
J(1, y, H) = -L(y_L - y)\mathbf{1}_{\{y < y_L\}} + H(y_H - y)\mathbf{1}_{\{y < y_H\}}.
$$

For $t < 1$ and $s \in \{L, 0, H\}$, set $J(t, y, s) = \mathsf{E}[J(t, Z_1, s) \,|\, Z_t = y]$. Then $J(t, Z_t, \xi S)$ is an $\mathbb{F}$–martingale, so it has zero drift. From Itô's formula, its drift is

$$
\frac{\partial}{\partial t} J(t, Z_t, \xi S) + \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial z^2} J(t, Z_t, \xi S).
$$

Equating this to zero, Itô's formula implies

$$
J(1, Y_1, \xi S) = J(0, 0, \xi S) + \int_0^1 \mathrm{d}J(t, Y_t, \xi S)
$$

$$
= J(0, 0, \xi S) + \int_0^1 \frac{\partial J(t, Y_t, \xi S)}{\partial y} \, \mathrm{d}Y_t .
$$

Therefore,

$$
\mathsf{E}[J(1, Y_1, \xi S) - J(0, 0, \xi S)] = \mathsf{E}\int_0^1 \frac{\partial J(t, Y_t, \xi S)}{\partial y} \, \mathrm{d}Y_t . \tag{A7}
$$

To calculate $\partial J(t, y, s)/\partial y$, use the fact that, by independent increments,

$$
J(t, y, s) = \mathsf{E}[J(t, Z_1, s) \,|\, Z_t = y] = \mathsf{E}[J(t, Z_1 - Z_t + y, s)]
$$

to obtain

$$
\frac{\partial J(t, y, s)}{\partial y} = \mathsf{E}\left[\frac{\partial}{\partial y} J(t, Z_1 - Z_t + y, s)\right].
$$

Now, note that, for any real number a excluding the kinks at $y_L - y$ and $y_H - y$,

$$\frac{\partial}{\partial y} J(1, a+y, L) = -L 1_{\{a > y_L - y\}} + H 1_{\{a > y_H - y\}},$$

$$\frac{\partial}{\partial y} J(1, a+y, 0) = L 1_{\{a < y_L - y\}} + H 1_{\{a > y_H - y\}},$$

$$\frac{\partial}{\partial y} J(1, a+y, H) = L 1_{\{a < y_L - y\}} - H 1_{\{a < y_H - y\}}.$$

Therefore,

$$\frac{\partial J(t, y, L)}{\partial y} = -L N\left(\frac{y - y_L}{\sigma\sqrt{1-t}}\right) + H N\left(\frac{y - y_H}{\sigma\sqrt{1-t}}\right),$$

$$\frac{\partial J(t, y, 0)}{\partial y} = L N\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) + H N\left(\frac{y - y_H}{\sigma\sqrt{1-t}}\right),$$

$$\frac{\partial J(t, y, H)}{\partial y} = L N\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) - H N\left(\frac{y_H - y}{\sigma\sqrt{1-t}}\right).$$

Now, the definition (6) gives us

$$\frac{\partial J(t, y, s)}{\partial y} = p(t, y) - s$$

for all $s \in \{L, 0, H\}$. Substituting this into (A7) gives us

$$\mathsf{E}[J(1, Y_1, \xi S) - J(0, 0, \xi S)] = \mathsf{E}\int_0^1 [p(t, Y_t) - \xi S]\mathrm{d}Y_t. \tag{A8}$$

The "no doubling strategies" condition implies that $\int p\,\mathrm{d}Z$ is a martingale, so the right-hand side of this equals

$$\mathsf{E}\int_0^1 [p(t, Y_t) - \xi S]\theta_t\,\mathrm{d}t.$$

Rearranging produces

$$\mathsf{E}\int_0^1 [\xi S - p(t, Y_t)]\theta_t\,\mathrm{d}t = \mathsf{E}[J(0, 0, \xi S) - J(1, Y_1, \xi S)] \leq \mathsf{E}[J(0, 0, \xi S)],$$

using the fact that $J(1, y, s) \geq 0$ for all (y, s) for the inequality. Thus, $\mathsf{E}[J(0, 0, \xi S)]$ is an upper bound on the expected profit, and the bound is achieved if and only if $J(1, Y_1, \xi S) = 0$ with probability one. By the definition of $J(1, y, s)$, this is equivalent to $Y_1 < y_L$ with probability one when $\xi S = L$, $y_L \leq Y_1 \leq y_H$ with probability one when $\xi = 0$, and $Y_1 > y_H$ with probability one when $\xi S = H$. By part (B) of the proposition, the strategy (5) is therefore optimal. ∎

Proof of Theorem 2

By Itô's formula and the fact that $(\mathrm{d}Y)^2 = (\mathrm{d}Z)^2 = \sigma^2\,\mathrm{d}t$, we have

$$\mathrm{d}p(t, Y_t) = \left(p_t(t, Y_t) + \frac{1}{2}\sigma^2 p_{yy}(t, Y_t)\right)\mathrm{d}t + p_y(t, Y_t)\mathrm{d}Y_t,$$

where we use subscripts to denote partial derivatives. Both Y and $p(t, Y_t)$ are martingales with respect to the market makers' information, so the drift term must be zero. That also can be verified by direct calculation of the partial derivatives, using the formula (6) for $p(t, y)$. Thus,

$$\mathrm{d}p(t, Y_t) = p_y(t, Y_t)\mathrm{d}Y_t.$$

A direct calculation based on the formula (6) for $p(t, y)$ shows that $p_y(t, y) = \lambda(t, y)$ defined in (7).

To see that $\lambda(t, Y_t)$ is a martingale for $t \in [0, 1)$, with respect to market makers' information, we can calculate, for $t < u < 1$,

$$\mathsf{E}[\lambda(u, Y_u) \mid Y_t = y] = -\frac{L}{\sigma\sqrt{1-u}} \cdot \int_{-\infty}^{\infty} \mathrm{n}\left(\frac{y_L - y'}{\sigma\sqrt{1-u}}\right) f(y' \mid u - t, y) \mathrm{d}y'$$

$$+ \frac{H}{\sigma\sqrt{1-u}} \cdot \int_{-\infty}^{\infty} \mathrm{n}\left(\frac{y_H - y'}{\sigma\sqrt{1-u}}\right) f(y' \mid u - t, y) \mathrm{d}y',$$

where $f(\cdot \mid \tau, y)$ denotes the normal density function with mean y and variance $\sigma^2 \tau$. A straightforward calculation shows that this equals $\lambda(t, y)$. For example, to evaluate the first term, use the fact that

$$\frac{1}{\sigma\sqrt{1-u}} \mathrm{n}\left(\frac{y_L - y'}{\sigma\sqrt{1-u}}\right) f(y' \mid u - t, y)$$

$$= \frac{1}{\sigma\sqrt{1-t}} \mathrm{n}\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right) \times \frac{1}{\sqrt{2\pi\sigma^2(1-u)(u-t)/(1-t)}}$$

$$\times \exp\left(-\left(\frac{1-t}{2(1-u)(u-t)\sigma^2}\right)\left(y' - \frac{(1-u)y + (u-t)y_L}{1-t}\right)^2\right),$$

which integrates to

$$\frac{1}{\sigma\sqrt{1-t}} \mathrm{n}\left(\frac{y_L - y}{\sigma\sqrt{1-t}}\right),$$

because the other factors constitute a normal density function. ∎

Appendix B. Hybrid Model Likelihood Function

Assume the trading period $[0, 1]$ corresponds to a day. This implies that any private information becomes public before trading opens on the following day.[28] We can estimate the model parameters using intraday price and order flow information. If we further assume that the model parameters are stable over time, then the price and order flow information from multiple days can be merged to estimate the parameters with greater precision.

To obtain stationarity in returns, assume that the possible signal realizations on each day are proportional to the observed opening price. Specifically, on each day i, assume that the possible signal realizations are

$$L_i = 2(p_L - 1)\kappa P_{i0},$$

$$H_i = 2p_L \kappa P_{i0},$$

where P_{i0} denotes the opening price on day i and where κ is a parameter to estimated. With this specification, the signal on each day has a zero mean, and $(H_i - L_i)/P_{i0} = 2\kappa$. Thus, κ measures the signal magnitude. Denote the pricing function on day i (as specified in Theorem 1) by $p_i(t, y)$, and let $p(t, y)$ denote the pricing function when the possible signal realizations are $L = 2(p_L - 1)\kappa$ and $H = 2p_L\kappa$. Then $p_i(t, y)/P_{i0} = p(t, y)$.

---

[28] In contrast to Odders-White and Ready (2008), our estimation does not use overnight returns. In our theoretical model, private information that is made public at the close of trading is incorporated into prices before trading ends (convergence to strong-form efficiency). Thus, overnight returns in our model are due to arrival of new public information, which does not aid in estimating the model.

The price at time t on day i is $V_{it} + p_i(t, Y_{it})$, and in particular the opening price is $P_{i0} = V_{i0}$, so the gross return through time t is

$$\frac{P_{it}}{P_{i0}} = \frac{V_{it}}{V_{i0}} + \frac{p_i(t, Y_{it})}{P_{i0}} = \frac{V_{it}}{V_{i0}} + p(t, Y_{it}). \tag{B1}$$

Assume

$$\frac{dV_{it}}{V_{it}} = \delta \, dB_{it}$$

for a constant δ and a Brownian motion B_i, so we have

$$\frac{P_{it}}{P_{i0}} = p(t, Y_{it}) + e^{\delta B_{it} - \delta^2 t/2}.$$

Assume the price and order imbalance are observed at times $t_1, \ldots, t_{k+1}$ each day with $t_{k+1} = 1$ being the close and the other times being equally spaced: $t_j = j\Delta$ for $\Delta > 0$ and $j \leq k$. Let P_{ij} denote the observed price and Y_{ij} the observed order imbalance at time t_j on date i. Let Γ denote the $(k+1)$–dimensional vector defined by $\Gamma_j = t_j/\Delta$ for $j = 1, \ldots, k+1$. Let Σ denote the $(k+1) \times (k+1)$ matrix defined by $\Sigma_{jj'} = \min(\Gamma_j, \Gamma_{j'})$.

Let U_i denote the vector of log pricing differences as defined in (10). The density function of $(P_{i1}/P_{i0}, \ldots, P_{i,k+1}/P_{i0})$ conditional on Y_i is

$$f(U_{i1}, \ldots U_{i,k+1}) e^{-\sum_{j=1}^{k+1} U_{ij}},$$

where f denotes the multivariate normal density function with mean vector $-(\delta^2 \Delta/2)\Gamma$ and covariance matrix $\delta^2 \Delta \Sigma$. Furthermore, on each day i, the vector $Y_i = (Y_{i,t_1}, \ldots, Y_{i,t_{k+1}})'$ is normally distributed with mean 0 and covariance matrix $\sigma^2 \Delta \Sigma$.

Let $\mathcal{L}_i$ denote the log-likelihood function for day i. Dropping terms that do not depend on the parameters, we have

$$-\mathcal{L}_i = (k+1)\log \sigma + \frac{1}{2\sigma^2 \Delta} Y_i' \Sigma^{-1} Y_i + (k+1)\log \delta$$

$$+ \frac{1}{2\delta^2 \Delta} \left(U_i + \frac{\delta^2 \Delta}{2} \Gamma \right)' \Sigma^{-1} \left(U_i + \frac{\delta^2 \Delta}{2} \Gamma \right) + \sum_{j=1}^{k+1} U_{ij}.$$

Using the facts that $\Gamma' \Sigma^{-1} = (0, \ldots, 0, 1)$ and $\Gamma' \Sigma^{-1} \Gamma = 1/\Delta$, this simplifies to

$$-\mathcal{L}_i = (k+1)\log \sigma + \frac{1}{2\sigma^2 \Delta} Y_i' \Sigma^{-1} Y_i + (k+1)\log \delta$$

$$+ \frac{1}{2\delta^2 \Delta} U_i' \Sigma^{-1} U_i + \frac{1}{2} U_{i,k+1} + \frac{\delta^2}{8} + \sum_{j=1}^{k+1} U_{ij}.$$

Hence, the log-likelihood function $\mathcal{L}$ for an observation period of n days satisfies (9).

References

Akay, O., K. B. Cyree, M. D. Griffiths, and D. B. Winters. 2012. What does PIN identify? Evidence from the T-bill market. *Journal of Financial Markets* 15:29–46.

Akins, B., J. Ng, and R. S. Verdi. 2012. Investor competition over information and the pricing of information asymmetry. *Accounting Review* 87:35–58.

Aktas, N., E. de Bodt, F. Declerck, and H. Van Oppens. 2007. The PIN anomaly around M&A announcements. *Journal of Financial Markets* 10:160–91.

Andersen, T., and O. Bondarenko. 2014a. Reflecting on the VPIN dispute. *Journal of Financial Markets* 17:53–64.

———. 2014b. VPIN and the flash crash. *Journal of Financial Markets* 17:1–46.

Back, K. 1992. Insider trading in continuous time. *Review of Financial Studies* 5:387–409.

Back, K., and S. Baruch. 2004. Information in securities markets: Kyle meets Glosten and Milgrom. *Econometrica* 72:433–65.

Banerjee, S., and B. Breon-Drish. 2017. Dynamic information acquisition and strategic trading. Working Paper, University of California, San Diego.

Banerjee, S., and B. Green. 2015. Signal or noise? Uncertainty and learning about whether other traders are informed. *Journal of Financial Economics* 117:398–423.

Benos, E., and M. Jochec. 2007. Testing the PIN variable. Working Paper, University of Illinois.

Bharath, S. T., P. Pasquariello, and G. Wu. 2009. Does asymmetric information drive capital structure decisions? *Review of Financial Studies* 22:3211–43.

Brennan, M. J., S.-W. Huh, and A. Subrahmanyam. 2016. High-frequency measures of informed trading and corporate announcements. Working Paper, UCLA and SUNY Buffalo.

Brown, S., and S. A. Hillegeist. 2007. How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies* 12:443–77.

Brown, S., S. A. Hillegeist, and K. Lo. 2004. Conference calls and information asymmetry. *Journal of Accounting and Economics* 37:343–66.

———. 2009. The effect of earnings surprises on information asymmetry. *Journal of Accounting and Economics* 47:208–25.

Chakraborty, A., and B. Yilmaz. 2004. Manipulation in market order models. *Journal of Financial Markets* 7:187–206.

Chen, Q., I. Goldstein, and W. Jiang. 2007. Price informativeness and investment sensitivity to stock price. *Review of Financial Studies* 20:619–50.

Collin-Dufresne, P., and V. Fos. 2015. Do prices reveal the presence of informed trading? *Journal of Finance* 70:1555–82.

Duarte, J., X. Han, J. Harford, and L. Young. 2008. Information asymmetry, information dissemination and the effect of Regulation FD on the cost of capital. *Journal of Financial Economics* 87:24–44.

Duarte, J., E. Hu, and L. Young. 2017. Does the PIN model identify private information and if so, what are our alternatives? Working Paper, Rice University, SEC, and University of Washington.

Duarte, J., and L. Young. 2009. Why is PIN priced? *Journal of Financial Economics* 91:119–38.

Easley, D., S. Hvidkjaer, and M. O'Hara. 2002. Is information risk a determinant of asset returns? *Journal of Finance* 57:2185–221.

———. 2010. Factoring information into returns. *Journal of Financial and Quantitative Analysis* 45:293–309.

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman. 1996. Liquidity, information, and infrequently traded stocks. *Journal of Finance* 51:1405–36.

Easley, D., M. López de Prado, and M. O'Hara. 2011. The microstructure of the "flash crash": Flow toxicity, liquidity crashes, and the probability of informed trading. *Journal of Portfolio Management* 37:118–28.

———. 2012. Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies* 25:1457–93.

———. 2014. VPIN and the flash crash: A rejoinder. *Journal of Financial Markets* 17:47–52.

Easley, D., and M. O'Hara. 2004. Information and the cost of capital. *Journal of Finance* 59:1553–83.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–636.

Ferreira, M. A., and P. A. Laux. 2007. Corporate governance, idiosyncratic risk, and information flow. *Journal of Finance* 62:951–89.

Foster, F. D., and S. Viswanathan. 1995. Can speculative trade explain the volume-volatility relation? *Journal of Business & Economic Statistics* 13:379–96.

Frankel, R., and X. Li. 2004. Characteristics of a firm's information environment and the information asymmetry between insiders and outsiders. *Journal of Accounting and Economics* 37:229–59.

Gan, Q., W. C. Wei, and D. Johnstone. 2017. Does the probability of informed trading model fit empirical data? *Financial Review* 52:5–35.

Glosten, L. R., and L. E. Harris. 1988. Estimating the components of the bid/ask spread. *Journal of Financial Economics* 21:123–42.

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

Goldstein, I., and A. Guembel. 2008. Manipulation and the allocational role of prices. *Review of Economic Studies* 75:133–64.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka. 2009. Do liquidity measures measure liquidity? *Journal of Financial Economics* 92:153–81.

Hasbrouck, J. 1988. Trades, quotes, inventories, and information. *Journal of Financial Economics* 22:229–52.

———. 1991. Measuring the information content of stock trades. *Journal of Finance* 46:179–207.

———. 2009. Trading costs and returns for U.S. equities: Estimating effective costs from daily data. *Journal of Finance* 64:1445–77.

Hendershott, T., and P. Moulton. 2011. Automation, speed, and stock market quality: The NYSE's hybrid. *Journal of Financial Markets* 14:568–604.

Holden, C. W., and S. Jacobsen. 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance* 69:1747–85.

Holden, C. W., and A. Subrahmanyam. 1992. Long-lived private information and imperfect competition. *Journal of Finance* 247–70.

Hwang, L. S., W. J. Lee, S. Y. Lim, and K. H. Park. 2013. Does information risk affect the implied cost of equity capital? An analysis of PIN and adjusted PIN. *Journal of Accounting and Economics* 55:148–67.

Jayaraman, S. 2008. Earnings volatility, cash flow volatility, and informed trading. *Journal of Accounting Research* 46:809–51.

Karatzas, I., and S. E. Shreve. 1988. *Brownian motion and stochastic calculus*. New York: Springer.

Kim, O., and R. E. Verrecchia. 1997. Pre-announcement and event-period private information. *Journal of Accounting and Economics* 24:394–419.

Korajczyk, R. A., and R. Sadka. (2008). Pricing the commonality across alternative measures of liquidity. *Journal of Financial Economics* 87:45–72.

Krinsky, I., and J. Lee. 1996. Earnings announcements and the components of the bid-ask spread. *Journal of Finance* 51:1523–35.

Kyle, A. S. 1985. Continuous auctions and insider trading. *Econometrica* 53:1315–36.

Lee, C. M., and M. J. Ready. 1991. Inferring trade direction from intraday data. *Journal of Finance* 46:733–46.

Li, H., J. Wang, C. Wu, and Y. He. 2009. Are liquidity and information risks priced in the Treasury bond market? *Journal of Finance* 64:467–503.

Mohanram, P., and S. Rajgopal. 2009. Is PIN priced risk? *Journal of Accounting and Economics* 47:226–43.

Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708.

Odders-White, E. R., and M. J. Ready. 2008. The probability and magnitude of information events. *Journal of Financial Economics* 87:227–48.

Rogers, L. C. G., and D. Williams. 2000. *Diffusions, Markov processes and martingales: Itô calculus*, vol. 2, 2nd ed. Cambridge: Cambridge University Press.

Rossi, S., and K. Tinn. 2010. Man or machine? Rational trading without information about fundamentals. Working Paper, Purdue University and Imperial College.

Venter, J. H., and D. de Jongh. 2006. Extending the EKOP model to estimate the probability of informed trading. *Studies in Economics and Econometrics* 30:25–39.

Wang, Y., and M. Yang. 2017. Insider trading when there may not be an insider. Working Paper, Duke University.