

Marketwide Private Information in Stocks: Forecasting Currency Returns

RUI ALBUQUERQUE, EVA DE FRANCISCO, and LUIS B. MARQUES*

ABSTRACT

We present a model of equity trading with informed and uninformed investors where informed investors trade on firm-specific and marketwide private information. The model is used to identify the component of order flow due to marketwide private information. Estimated trades driven by marketwide private information display little or no correlation with the first principal component in order flow. Indeed, we find that co-movement in order flow captures variation mostly in liquidity trades. Marketwide private information obtained from equity market data forecasts industry stock returns, and also currency returns.

MARKETS AGGREGATE DISPERSED INFORMATION from economic agents and impound it into prices. This information originates from public sources, such as company reports and official statistics, or from the proprietary models, expertise, and insider knowledge of private investors. Its content is either asset-specific or aggregate. By definition, *marketwide private information* is useful for trading across a variety of assets. In the context of stock trading, marketwide private information can be informative about future firm cash flows that fluctuate with industry or economy-wide business conditions, or about discount rates that move with the economy's riskless interest rate and aggregate risk premium. In contrast, *firm-specific private information* is idiosyncratic and useless for the valuation of other stocks or assets. While much has been written about firm-specific private information, not much is known about marketwide private information.

This paper contributes to the study of marketwide private information in two ways. First, we construct a model of stock trading with informed and uninformed investors to structurally identify marketwide private information from firm-specific private information and from liquidity trades. The model generalizes Easley et al. (hereafter EKOP (1996)) by allowing for trading in multiple

*Albuquerque is at the Boston University School of Management and with CEPR. De Francisco is at the Department of Economics at Towson University. Marques is at the Paul H. Nitze School of Advanced International Studies, Johns Hopkins University and with CEMPRE. We thank Marine Carrasco and Werner Ploberger for useful discussions and Clara Vega for many detailed comments on an earlier version of the paper. We also thank seminar participants at the Bank of Portugal, Instituto de Empresa Business School in Madrid, the University of Rochester and the Federal Reserve Board for comments. The paper has greatly benefited from comments by an anonymous referee and an associate editor. Marques thanks Fundação para a Ciência e a Tecnologia, Portugal, for financial support. The usual disclaimer applies.

stocks and in private information at two levels, the firm-specific level and the marketwide level. Our identifying assumption is that marketwide private information generates trading simultaneously in several stocks. Good (bad) marketwide private information generates informed investor-initiated buy (sell) orders across all firms. In contrast, good (bad) firm-specific private information leads to increased informed investor-initiated buy (sell) orders in that firm alone. The possibility that marketwide private information and firm-specific private information will offset each other requires an *ex ante* choice of weights to each information signal. We assume that informed investors discard their marketwide private signals, which is consistent with the view that firm-specific private information is generally more precise. The model is estimated by maximum likelihood using microstructure stock trading data from five industries.¹ We use the parameter estimates to construct measures of monthly industry order flow due to marketwide private information. Goodness of fit tests are performed to validate the estimation results.

The paper's second contribution is the study of properties of marketwide private information. Estimated marketwide private information has three main properties. First, in each industry that we study marketwide private information displays little or no correlation with total order flow or with the first principal component taken from the order flow of the firms in the industry. The latter fact implies that marketwide private information cannot be replicated by a simple statistical procedure, such as a principal component analysis. Instead, our estimate of liquidity trades correlates strongly with the principal component in order flow, providing an interpretation for the co-movement in order flow observed in Hasbrouck and Seppi (2001). Moreover, because most of the firms we study are not index constituents, our analysis is not subject to Harford and Kaul's (2005) criticism that co-movement in order flow is driven by firms sharing the same index.

The second property of marketwide private information is the ability to forecast the stock returns of the firms in the industries we study. Marketwide private information is able to forecast returns up to 2 months ahead, except for the industry with the fewest firms needed to estimate the model. This finding suggests that we are indeed capturing information-driven trading as opposed to noninformative inventory or liquidity effects.

The third property of marketwide private information is that it forecasts currency returns. This finding gives additional credence to our label of marketwide private information and constitutes evidence in favor of Evans and Lyons (2004a). We regress changes in industry-specific currency baskets on marketwide private information. We repeat the same regressions using the

¹ These industries are chosen to satisfy two main criteria. First, these industries have high average ratios of exports relative to total shipments. Our choice is aimed at finding industries for which marketwide private information, if it exists, is most likely correlated with factors that also drive exchange rates. Firms in our industries are shown to have qualitatively similar foreign currency exposures. The second main selection criterion is that these industries have multiple firms trading in a liquid fashion in the NYSE. Appendices A and B contain the details.

main currencies that compose each industry-specific basket. We use as dependent variables the simple currency return (i.e., percentage change of the exchange rate) and the excess currency return (i.e., percentage change of the exchange rate minus the interest rate differential). Our measures of marketwide private information forecast current currency returns, 1- and 2-months ahead currency returns, and excess currency returns of the industry-specific baskets with R^2 s between 5% and 16%. The measures of marketwide private information can also forecast the main currencies that compose each basket, in some cases displaying R^2 s up to 23%.

Evans and Lyons (2004a) assume the existence of marketwide private information and present a model that can explain the contemporaneous correlation between exchange rate returns and currency order flow observed by Evans and Lyons (2002) and Rime (2001). Their model also predicts that equity market order flow driven by marketwide private information forecasts exchange rate returns. Our findings provide direct evidence consistent with their model: (i) marketwide private information exists; and (ii) marketwide private information permeates both equity and currency markets.

Our findings also help validate asset pricing and microstructure models that start with the premise that marketwide private information exists. In international finance, two prominent examples are Gehrig (1993) and Brennan and Cao (1997), who respectively explain the home bias and return chasing of U.S. investors. These papers implicitly hypothesize the existence of asymmetric information about stock market country *indices*: If all private information were about firm-specific factors, then one would not expect most of it to survive aggregation in large, diversified country indices where these factors are absent. Bacchetta and van Wincoop (2006) postulate that agents have aggregate private information that is used in both the money market and the foreign exchange market. Albuquerque, Bauer, and Schneider (2005) hypothesize that U.S. investors have better private information about global factors than do local investors and show that this implies a pattern of global return chasing. In the market microstructure literature, Chan (1993) demonstrates that if marketwide private information exists, then market makers observing only signals on their own stock generate pricing errors that correlate with those in other stocks, leading to positive cross-autocorrelation in returns, a phenomenon pervasive in the data. Subrahmanyam (1991) shows that trading in stock indices is advantageous to discretionary liquidity traders even in the presence of investors who are informed about systematic factors. Kumar and Seppi (1994) develop a model of index arbitrage in which informed investors receive private signals about the only factor driving cash flows. Finally, Caballe and Krishnan (1994) build a general multi-agent, multi-security asset pricing framework with marketwide private information.

In spite of marketwide private information's prominent role in asset pricing theories, little evidence has been presented in favor of its existence. This paper is the first to provide a measure of marketwide private information obtained

from the estimation of a structural model of asset trading. In the past, studies have discussed the existence of marketwide private information via either its indirect effects or the estimation of nonstructural models. Barclay, Litzenberger, and Warner (1990) study the relation between index return volatility and trading volume and find evidence consistent with the existence of marketwide private information in the Tokyo stock exchange, but do not directly test for it. Albuquerque et al. (2005) find a common factor in private information from aggregate net-purchases data of U.S. investors on eight developed country equity markets. Their measure is a statistical decomposition of flows that reflects covariation in unexpected net flows of U.S. investors across these markets (see also Bauer and Vega (2004) and Yu (2005)). Chan and Hameed (2006) show that firms with greater analyst coverage display greater synchronicity with the market and interpret synchronicity as having prices that incorporate greater marketwide (private *and* public) information.

Our results complement those of Evans and Lyons (2002, 2004b) and Rime (2001), who link currency order flow with currency returns,² and those of the domestic microstructure literature (e.g., Hasbrouck (1991)) that links stock order flow with own-stock returns. In our paper, marketwide private information links order flow in the stock market, driven by marketwide private information, to industry stock returns and currency returns. Our results are also complementary to those in Francis, Hasan, and Hunter (2006), who study the reverse information spillover of currency order flow into the equity market and find mostly effects in volatilities.

The remainder of the paper is organized as follows. In Section I, we develop a theoretical model of trading that allows us to estimate a measure of marketwide private information and to conduct our hypothesis tests. In Section II, we provide details on the data. Section III presents results on the estimation of marketwide private information and on tests of the main hypotheses. Section IV concludes the paper. The Appendix gives additional details on the sample selection, on the currency exposure of the firms in our sample, and on properties of the model estimates.

I. The Model of Stock Trading

This section presents a model of trading that allows for firm-specific and marketwide private information. The goal of this section is to identify through a structural model the component of observed order flow that is due to marketwide private information from all else.

A. Trading

The model is one of sequential trading where informed and uninformed investors post buy and sell orders to a market maker. All agents are risk neutral

² Evans and Lyons (2004b) show that order flow in foreign currency markets can also forecast output growth, money growth, and inflation.

and competitive. Following EKOP, we assume that traders trade during a finite number of days, but that trading in each day is continuous and the arrival of uninformed and informed traders is determined by independent Poisson processes. The market maker chooses prices consistent with the observed aggregate order flow from informed and uninformed investors. However, as shown in EKOP and Easley and O'Hara (1992), there is no explicit feedback from prices to order flow. Hence, prices play no role in our analysis of order flow and their presentation is omitted from the paper. We deviate from EKOP by allowing investors and the market maker to trade on $I > 1$ risky stocks indexed by $i = 1, \dots, I$ and cash balances.

Prior to the start of any trading day, informed investors may receive up to $I + 1$ bits of information. With probability θ an information event occurs containing marketwide private information that is useful across all I assets. Such an event contains good (bad) marketwide private information news with probability $1 - \rho(\rho)$ and affects positively (negatively) all I assets, though not necessarily in the same magnitude. For firm i , with probability α_i an information event occurs containing firm-specific private information useful only for firm i . Such an event contains good (bad) private information news with probability $1 - \delta_i(\delta_i)$. The arrival of marketwide private information is independent of the arrival of firm-specific private information; marketwide and firm-specific private information on the same event are possible, but are assumed to arrive independently. The full information value of each asset from the previous day is revealed before trading starts.

Figure 1 describes the information available on firm i and the trading activity that is generated on average at each node. Days can have both marketwide private information and firm-specific private information, only the former, only the latter, or neither. With more than one piece of news affecting trading we adopt a simple rule for conflict resolution: Whenever marketwide and firm-specific private information on any firm i are qualitatively contradictory, the firm-specific news dominates investors' behavior, which is consistent with the view that marketwide private information is generally composed of less precise information. This assumption is reflected in the absence of μ_i^a (see below) in lines 4 and 7 on the right-hand side of Figure 1. Informed investors may therefore not act in accordance with their marketwide private information.

Consider, for example, a day with good marketwide private information. The day has three possible outcomes. If the day has no firm-specific information event for firm i , overall news on firm i is good. If the day results in good firm-specific private information, then the two bits of information are reinforcing and overall news is good. Finally, if the day has bad firm-specific private information, then the two bits of information are contradictory and overall news is bad.

At the end of each node, marked with a square in the figure, the trading day starts and trades arrive continuously and independently according to known Poisson processes. For firm i , let μ_i^f be the average arrival rate of informed investors who trade based on firm-specific private information news and let μ_i^a be

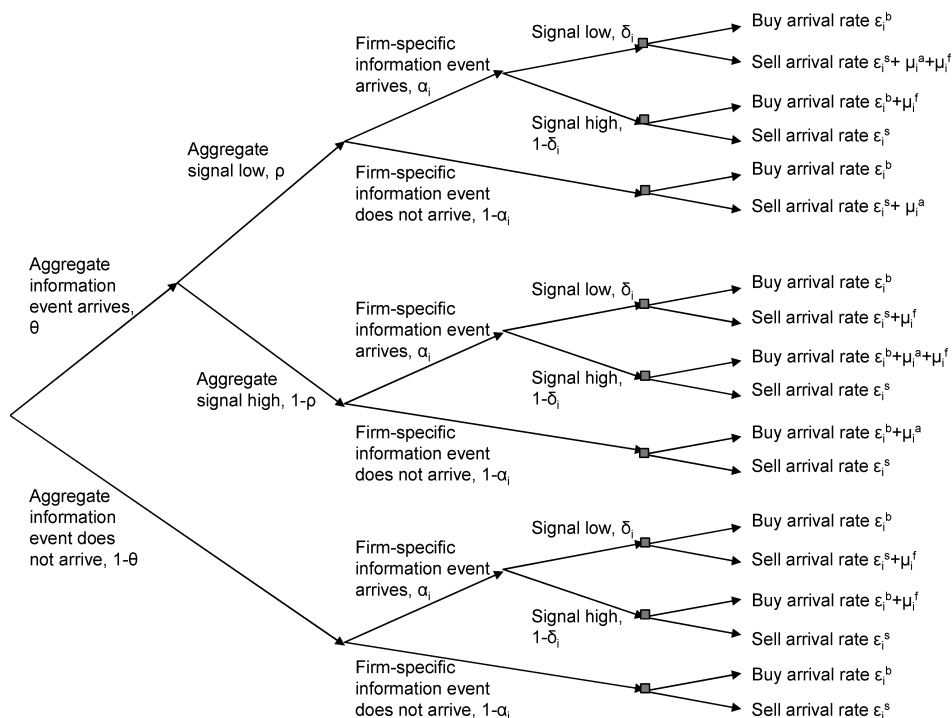


Figure 1. Tree diagram of the trading process for stock i . θ is the probability of a marketwide information event, p is the probability of a marketwide low signal, α_i is the probability of a firm i -specific information event, δ_i is the probability of a low signal for firm i , $\epsilon_i^b(\epsilon_i^s)$ is the mean arrival rate of buy (sell) orders from uninformed investors to firm i , and $\mu_i^a(\mu_i^f)$ is the mean arrival rate of trades by informed investors based on marketwide (firm-specific) private signals.

the average arrival rate of informed investors who trade based on marketwide private information news. The index i on both μ_i^f and μ_i^a means that each firm is allowed different sensitivities to firm-specific and marketwide factors. Uninformed investors' buy orders arrive at an average rate ϵ_i^b and sell orders arrive at an average rate ϵ_i^s . These parameters are constant and known to everyone for the duration of the trading period, but the market maker does not know if he is trading against an informed or an uninformed investor. With this notation, consider the node at the top of Figure 1 for firm i obtained on a day with bad marketwide and firm-specific private information news. The total volume of sell orders has an average of $\mu_i^f + \mu_i^a + \epsilon_i^s$, whereas only uninformed investors buy, making the average total volume of buy orders ϵ_i^b . At the bottom of the tree, the event with no marketwide private information and no firm-specific private information on firm i generates trading only by uninformed investors. The average number of buy orders is ϵ_i^b and that of sell orders is ϵ_i^s . The rest of the arrival rates at the end of each node are constructed in a similar fashion.

B. The Likelihood Function

We now construct the likelihood function of observing $\{(S_{in}, B_{in})_{n=1}^N\}_{i=1}^I$ sell and buy orders, across I firms during N trading days. A day n with good marketwide private information news occurs with probability $\theta(1 - \rho)$. Given good marketwide private information, the conditional probability of observing the pair of sell and buy orders $\{S_i, B_i\}$ for firm i is

$$\begin{aligned} l_G(\{S_i, B_i\}) = & \alpha_i(1 - \delta_i)e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-(\mu_i^f + \mu_i^a + \varepsilon_i^b)} \frac{(\mu_i^f + \mu_i^a + \varepsilon_i^b)^{B_{in}}}{B_{in}!} \\ & + \alpha_i \delta_i e^{-(\mu_i^f + \varepsilon_i^s)} \frac{(\mu_i^f + \varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!} \\ & + (1 - \alpha_i) e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-(\mu_i^a + \varepsilon_i^b)} \frac{(\mu_i^a + \varepsilon_i^b)^{B_{in}}}{B_{in}!}. \end{aligned} \quad (1)$$

Under the assumption of the independence of buy and sell orders across firms, the probability of observing $\{S_{in}, B_{in}\}_{i=1, \dots, I}$ on day n of good marketwide private news is $\prod_{i=1}^I l_G(\{S_{in}, B_{in}\})$.

A day n with bad marketwide private information news occurs with probability $\theta\rho$. Given such marketwide private information, the conditional probability of observing the pair of sell and buy orders $\{S_i, B_i\}$ for firm i is

$$\begin{aligned} l_B(\{S_i, B_i\}) = & \alpha_i(1 - \delta_i) e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-(\mu_i^f + \varepsilon_i^b)} \frac{(\mu_i^f + \varepsilon_i^b)^{B_{in}}}{B_{in}!} \\ & + \alpha_i \delta_i e^{-(\mu_i^f + \mu_i^a + \varepsilon_i^s)} \frac{(\mu_i^f + \mu_i^a + \varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!} \\ & + (1 - \alpha_i) e^{-(\mu_i^a + \varepsilon_i^s)} \frac{(\mu_i^a + \varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!}, \end{aligned} \quad (2)$$

and the probability of observing the buy and sell orders $\{S_{in}, B_{in}\}_{i=1, \dots, I}$ on day n of bad marketwide private news is $\prod_{i=1}^I l_B(\{S_{in}, B_{in}\})$.

Finally, a day with no marketwide private information news occurs with probability $1 - \theta$, and on these days the probability of observing the pair of sell and buy orders $\{S_i, B_i\}$ for firm i is

$$\begin{aligned} l_0(\{S_i, B_i\}) = & \alpha_i(1 - \delta_i) e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-(\mu_i^f + \varepsilon_i^b)} \frac{(\mu_i^f + \varepsilon_i^b)^{B_{in}}}{B_{in}!} \\ & + \alpha_i \delta_i e^{-(\mu_i^f + \varepsilon_i^s)} \frac{(\mu_i^f + \varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!} \\ & + (1 - \alpha_i) e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!}. \end{aligned} \quad (3)$$

The probability of observing the buy and sell orders $\{S_{in}, B_{in}\}_{i=1, \dots, I}$ on day n of no marketwide private news is $\Pi_{i=1}^I l_0(\{S_{in}, B_{in}\})$.

On any day n , the unconditional likelihood of observing I buy orders $\{B_{in}\}$ and I sell orders $\{S_{in}\}$ is the weighted average of the expressions above, with the weights given by the probability of each type of marketwide information event:

$$l((S_{in}, B_{in})_{i=1, \dots, I}) = \theta(1 - \rho)\Pi_{i=1}^I l_G(\{S_{in}, B_{in}\}) + \theta\rho\Pi_{i=1}^I l_B(\{S_{in}, B_{in}\}) + (1 - \theta)\Pi_{i=1}^I l_0(\{S_{in}, B_{in}\}). \quad (4)$$

The likelihood of observing $I \times N$ buy orders $\{B_{in}\}$ and $I \times N$ sell orders $\{S_{in}\}$ is then

$$L((S_{in}, B_{in})_{n=1, \dots, N; i=1, \dots, I}) = \Pi_{n=1}^N l((S_{in}, B_{in})_{i=1, \dots, I}). \quad (5)$$

The likelihood function (5) is maximized to solve for $(\alpha_i, \delta_i, \theta, \rho, \mu_i^a, \mu_i^f, \varepsilon_i^s, \varepsilon_i^b)_{i=1, \dots, I}$, where all parameters except for θ and ρ vary by firm. Because this problem does not admit a closed-form solution, we resort to numerical methods to estimate the model (see Section III).

Allowing the parameters μ_i^f , μ_i^a , ε_i^b , and ε_i^s to vary with i gives the model flexibility to capture different trading intensities to news across firms. To estimate the parameters that drive the release of firm-specific private information (α_i, δ_i) , we need the existence for each firm of common time-series patterns in order flow. To see why this is the case, note that conditional on marketwide private information, daily likelihoods are trinomials of Poisson probability functions, which are bilinear in α_i and δ_i (see Easley, Kiefer, and O'Hara (1997) and Vega (2006)). In the absence of informed trading, $(\varepsilon_i^s, \varepsilon_i^b)$ measure the average number of sell orders and the average number of buy orders for firm i , respectively (see footnote 10 below). With informed trading, $(\varepsilon_i^s, \varepsilon_i^b)$ still capture average trading, while μ_i^f measures the abnormal number of buy or sell orders that are firm-specific (Vega (2006)). This is also true in our model with the exception of some qualifications that we address below.

In our model μ_i^a captures abnormal trading, but only the abnormal trading that occurs consistently across firms in a day. The richness of possible events in the model permits estimation of μ_i^a separately from μ_i^f : When private marketwide and firm-specific news agree, informed trading is abnormally higher than when they do not. Loosely speaking, our model uses the time-series fluctuations in average (across firms) order flow to identify the common parameters (θ, ρ) , with days characterized by common patterns across firms pushing estimates of θ up and, of these days, those characterized by common high levels of sell (buy) orders pushing estimates of ρ up (down).

In contrast to Easley et al. (1997) and Vega (2006), trades posted to each firm are not independent, meaning that informed investors have aggregate news useful to trade across all firms. This implies that the estimation of firm i 's parameters α_i , δ_i , μ_i^a , μ_i^f , ε_i^s , and ε_i^b depends on the estimation of the other firms' parameters because they are linked by the arrival of marketwide private

information news. To see this connection consider solving a maximum likelihood problem as in Easley et al. (1997) and Vega (2006), who study every firm in isolation when the true model is one with marketwide private information. Biases in the estimation of α_i , ε_i^s , and ε_i^b can occur, for example, if days are perceived by such optimization as having no firm-specific private information news when they are actually days with marketwide private information. This is more likely to occur if the true μ_i^a is sufficiently small, leading to an upward bias in $\hat{\varepsilon}_i^s$ and $\hat{\varepsilon}_i^b$.³ If, in contrast, the true μ_i^a is close to the true μ_i^f , a day with marketwide private information will be mistaken for a day with firm-specific private information, biasing the value of $\hat{\mu}_i^f$ downwards (upwards) if $\mu_i^f > \mu_i^a$ ($<$) and affecting the inference of α_i and δ_i as well. It is easy to construct more examples, but they are not instructive, as there is no simple way of describing exactly how each parameter is estimated in the maximization of the likelihood function (i.e., mathematically, the first-order conditions are nonlinear and do not admit a closed-form solution).

The structural model developed here allows some heterogeneity in firm responses to marketwide factors. Specifically, as indicated above, firms may have different intensities of trading due to marketwide private information, μ_i^a . By capturing the marketwide shocks as firm-specific events for the firms negatively affected, the model also allows some firms to be negatively affected by shocks to a marketwide factor, while the majority of firms are positively affected by the same shock (or vice versa). However, this result requires such marketwide shocks to be infrequent events; the model assumes that firm-specific and marketwide news are independent. If these events are frequent, then the estimation may still capture them via the trading activity of the majority of firms, but will likely introduce a downward bias and an estimation error on estimates of marketwide private information. To minimize this bias, we have designed sample selection criteria that identify industries in which firms face significant exposure to currency risk and are affected in similar ways by that same risk. Section II.A and Appendices A and B contain the details.⁴

C. Decomposing Order Flow

With the estimated parameters one can construct an artificial measure of the average number of buy and sell orders in any one day that are induced by

³ The optimization thus assigns to liquidity trading the values $\varepsilon_i^s + \mu_i^a$ and $\varepsilon_i^b + \mu_i^a$ for those days.

⁴ Alternatively, this type of heterogeneity can be explicitly modeled, with the advantage of being potentially applicable to any industry. The drawbacks consist of increased computational cost and loss of degrees of freedom. Suppose a single source of marketwide news exists. Without taking an a priori stance on the model of heterogeneity, one has to allow for all the possible combinations of firms that respond positively versus those that respond negatively to such news. This has a significant computational cost, even if the number of parameters stays constant, given that each estimation we conduct requires an exhaustive grid search of initial conditions. With I firms in each industry, we multiply the number of estimations by 2^I (currently we estimate the model T months, but in this approach we need $T2^I$ estimations). The alternative to considering all possible combinations of heterogenous responses is to rely on ex ante industry information, which brings us back to our approach.

marketwide private information news. Informed investors are buyers of firm i based on marketwide private information when they hold good marketwide private information (which occurs with probability $\theta(1 - \rho)$) and if they have either no firm-specific private information on firm i (with probability $1 - \alpha_i$) or have good firm-specific private information (with probability $\alpha_i(1 - \delta_i)$):

$$\text{Average Marketwide-Informed Buys} = \theta(1 - \rho) \sum_{i=1}^I [1 - \alpha_i + \alpha_i(1 - \delta_i)] \mu_i^a. \quad (6)$$

In contrast, informed investors are sellers of firm i when they hold bad marketwide private information (which occurs with probability $\theta\rho$) and if they have either no firm-specific private information (with probability $1 - \alpha_i$) or have bad firm-specific private information (with probability $\alpha_i\delta_i$):

$$\text{Average Marketwide-Informed Sells} = \theta\rho \sum_{i=1}^I (1 - \alpha_i + \alpha_i\delta_i) \mu_i^a. \quad (7)$$

Combining (6) and (7), the industry daily average (across I) order flow driven by marketwide private information news is given by

$$MPI = \theta(1 - \rho) \sum_{i=1}^I [1 - \alpha_i + \alpha_i(1 - \delta_i)] \mu_i^a - \theta\rho \sum_{i=1}^I (1 - \alpha_i + \alpha_i\delta_i) \mu_i^a. \quad (8)$$

The variable MPI captures the qualitative nature of the marketwide private information embedded in the trades of investors in the industry. A positive MPI means that the industry was dominated by good aggregate news during the time period used for the estimation of the parameters, whereas a negative MPI implies the dominance of bad marketwide private information news. EKOP use the probability of informed trading, PIN , as a measure of private information. Note that PIN does not distinguish between good news days and bad news days, but instead distinguishes days with high levels of private information trading (sell or buy) from days with low levels of private information trading (sell or buy). This approach is appropriate for forecasting absolute returns (e.g., to study the speed of information diffusion) or if the focus is on the quantity of information asymmetry (e.g., EKOP), but PIN cannot be used to forecast actual returns.

D. MPI and Stock and Currency Returns

This subsection describes two hypothesis tests that serve as applications for our measure of marketwide private information. The first hypothesis follows Hasbrouck (1991) in identifying information shocks as those with a permanent impact on trades. Because marketwide private information MPI is such an information shock, it must forecast the returns of the firms in the industry from which it was obtained. Therefore, the first hypothesis we test is

Hypothesis 1: *MPI forecasts the equity returns of the firms in the industry from which it was obtained.*

In light of Hasbrouck's identification scheme, finding evidence in favor of Hypothesis 1 is a necessary condition to claiming that *MPI* is a measure of private information. Hypothesis 1 also tests a basic assumption of the model in Evans and Lyons (2004a) and of other models that rely on marketwide private information, as discussed in the introduction.

The second hypothesis we test is also related to Evans and Lyons (2004a). Evans and Lyons present a general equilibrium framework that explains the correlation between contemporaneous order flow in the foreign exchange market and currency returns. In their model, the presence of transitory and persistent productivity shocks means that flows into the equity market and flows into the currency market are correlated with and forecast changes in the exchange rate.⁵ This follows because investors trade based on their private information but set prices based only on public information. After trading, each investor's private information becomes public, which can be inferred without noise from order flow. Subsequently, exchange rate quotes are revised.

In Evans and Lyons (2004a) marketwide private information can only forecast one-period ahead currency returns; afterwards, the information contained in order flow is observed and immediately incorporated into prices and agents' trading strategies. In real economies, however, order flow contains noninformative liquidity trades, making marketwide private information trades not directly observable. It is thus likely that information contained in order flow diffuses slowly over time and remains useful to forecast currency returns further into the future (e.g., Evans and Lyons (2004b) for evidence of the slow adjustment of currency returns to information contained in currency transaction flow). This implies that lags of *MPI* may be informative about future returns above and beyond the information content of current *MPI*. We therefore obtain the following hypothesis:

Hypothesis 2: *Equity order flow driven by marketwide private information, MPI, forecasts changes in exchange rate returns.*

Evidence consistent with Hypothesis 2 is also consistent with *MPI*'s role as a measure of marketwide information, that is, *MPI* contains information that is relevant in a variety of markets.

In testing Hypothesis 2 we are not interested in all trades resulting from private information, but rather those trades that come from marketwide private information. This is because only the latter will have relevant information about aggregate factors that also drive exchange rates, as opposed to information about idiosyncratic factors that do not survive aggregation. To this end we use the above asset trading model that allows the identification of equity

⁵ Evans and Lyons (2004b) document that foreign exchange order flow has predictive power for aggregate variables including the exchange rate, but they do not establish the connection to the stock market as their model implies and as we do here.

market trades driven by marketwide private information from trades due to firm-specific private information and liquidity trades.

II. Data

In this section we first describe the industry and firm selection procedure, the data used, and its sources. Finally, we give details about the foreign currency exposure of the firms in our sample.

A. Industry and Firm Selection

The above method to extract marketwide private information is quite general and can be applied to any industry. However, given our objective of obtaining a measure of marketwide private information that is relevant to forecast currency returns, we choose to focus on industries with significant and common international exposure. Our goal is to increase the statistical power of our tests. A complete description of the sample selection is given in Appendix A.

To ensure homogeneity of firms in their foreign currency exposure we use the highest (six-digit) level of disaggregation per industry of the North American Industrial Classification System (NAICS). We measure international exposure using segment data on export sales relative to total shipments. Export sales data come from the U.S. International Trade Commission database and shipment data are from the Annual Survey of Manufactures of the U.S. Census Bureau. We start with the industries that rank in the top 30 in exports to total shipments in each year from 1997 to 2003, and keep only those that continuously ranked among the top 30. This leaves us with 20 industries. We then drop 11 industries that do not have a complete bridge between NAICS and its predecessor, the Standard Industrial Classification code (SIC).⁶ This guarantees that firms are treated consistently over the period from January 1993 to December 2003. We make use of international trade data to determine which currencies are most important for the selected industries and to construct a currency basket that gives an index of foreign exposure to the industry as a whole.

We now turn to the selection of firms within each of the remaining nine industries. We exclude firms that are not traded in the NYSE, firms with low market liquidity, and foreign firms. The requirement that firms must be traded in the NYSE is justified because the trading model we estimate applies best to the market-making trading environment of the NYSE. The liquidity requirement is applied month-by-month to each firm to guarantee a minimum of seven trades on average per day in any given month (EKOP). Firms that fail to meet the liquidity requirement on any given month are excluded from the estimation

⁶ In 1997, the Office of Management and Budget adopted NAICS, a system for classifying establishments by type of economic activity, to replace the 1987 Standard Industrial Classification. NAICS is constructed within a single production-oriented or supply-based conceptual framework and provides comparability in statistics about business activity across North America. The system was revised in 2002 but remained mostly unchanged for the manufacturing sector.

in that month. With the purpose of identifying firms with qualitatively similar exposures to exchange rates, we also exclude foreign firms that did not have a significant presence in terms of operations in the U.S. For example, this criterion leads us to include the U.K. firm Doncasters PLC in the Aircraft Engines Manufacturing industry and to exclude the Brazilian firm Embraer from the Aircraft Manufacturing industry.

To complete our selection procedure we restrict the set of possible industries to industries with at least four firms. The requirement for having several firms in an industry is justified by the following: (i) the need to identify those parameters that are firm-independent, that is, to identify the probability of marketwide private information events, θ , and the fraction of time marketwide news is bad news, ρ ; and (ii) the number of degrees of freedom increases with the number of firms (see below). This leaves us with seven industries. We drop one industry for having too many firms to permit estimation of the model and another due to low dollar export volume. The firms in the remaining five industries all have data available from both the Trade and Quotes (TAQ) database and the CRSP/Compustat database.

In our sample period between January 1993 and December 2003 some companies enter and others exit the sample, that is, the industry. The reasons for entry include initiation of trading in the NYSE or re-classification within NAICS (perhaps because of a merger or simply a change in business strategy). Similarly, the reasons for exit include bankruptcy, termination of trading in the NYSE, or a change in main business activity causing the firm to drop from the NAICS. In very few instances we are not able to avoid months for which only one company was traded while simultaneously satisfying the minimum liquidity criterion. For these months firm-specific private information order flow is equated to marketwide private information order flow.

B. Variable Definitions

Stock order flow is obtained from the TAQ database from January of 1993 to December of 2003. We use the Lee and Ready (1991) algorithm to calculate the daily number of buys and sells for each firm in our sample. Lee and Ready use the quote method to classify transactions whenever possible, labeling an order as a buy if the transaction price is above the spread midpoint and as a sell if the transaction price is below the spread midpoint, and leaving unclassified transactions at the spread midpoint. For the unclassified transactions, Lee and Ready use the tick method. The tick method classifies transactions by comparing the price of the current trade to the price of the preceding trade. Upticks (price increases relative to the previous transaction price) are buys, and downticks are sells. Zero-upticks (zero price changes such that the last price change was an uptick) are buys and zero-downticks are sells. Lee and Ready also argue that updated quotes are usually reported before the transactions that triggered them, implying that a comparison of the execution price to the quote in effect at the time of the transaction is inappropriate. They propose a five-second rule for comparing execution prices to quotes reported a minimum

of five seconds before the transaction was reported. Existing evidence suggests that the Lee and Ready algorithm is a good method for identifying the direction of trade from data in the TAQ database.⁷

Stock returns are defined as monthly holding period returns as provided in the CRSP database (from month-end to month-end with dividends reinvested at month-end). Exchange rate and interest rate data are taken from Datastream and are complemented with data from the International Financial Statistics of the International Monetary Fund. We use *beginning of month quotes* of foreign currency (FC) per U.S. dollar (\$ or USD), denoted by $S_{FC/\$t}$. Currency returns in month t , cr_t , are given by

$$cr_t \equiv \ln S_{FC/\$,t+1} - \ln S_{FC/\$,t}. \quad (9)$$

Excess currency returns in month t , xcr_t , are given by

$$xcr_t \equiv \ln S_{FC/\$,t+1} - \ln S_{FC/\$,t} - \ln(1 + i_{\$,t}) + \ln(1 + i_{FC,t}), \quad (10)$$

where $i_{FC,t}$ and $i_{\$,t}$ are the beginning of month t nominal interest rates on the FC and USD, respectively. A positive xcr represents an appreciation of the USD relative to the FC over and above a predicted change in exchange rates from the interest rate differential.

For each industry, we also obtain trade-weighted currency returns and excess currency returns using as weights the previous month's fraction of industry exports going to each country. For each month we allow at most five currencies in each currency basket, but these currencies can vary from year to year, according to the export weight of the corresponding countries. Trade weights are selected based on export performance in the immediately preceding year. Export weights were used in all industries except for NAICS 333132 where import weights were used (see below). Data for imports are customs value imports for consumption from the U.S. Trade Commission database.

C. The Industries under the Microscope

This subsection describes each of the five industries in the paper. In particular, it documents their foreign currency exposure and gives details on the industry trade-weighted currency return. Foreign currency exposure is generally associated with translation of local currency balances of foreign subsidiaries, intercompany loans with subsidiaries, transactions denominated in foreign currency, and operating exposure. While we discuss currency exposure we note that

⁷ Odders-White (2000) studies the performance of the Lee and Ready method using TORQ data and finds that it correctly classifies 85% of the transactions in her sample. Lee and Radhakrishna (2000) show that batched orders, stopped orders, and market crosses all add noise to the inference process. However, despite these problems, their results suggest that trades and quotes data provide substantial information about the original orders. Specifically, the active side of each trade, as identified by the Lee and Ready method, is generally a good proxy for the frequency, size, and direction of incoming market orders. Ellis, Michaely, and O'Hara (2000) used a Nasdaq proprietary data set that identifies trade direction to examine the accuracy of several trade classification algorithms. The Lee and Ready algorithm shows the best results, correctly classifying 81.05% of the trades.

many U.S. firms invoice in USD (e.g., Goldberg and Tille (2004)), a practice we also document for some of our firms. This does not mean that such firms are hedged against currency fluctuations; a USD depreciation, while not changing the unit price of exports, increases the demand for U.S. products (operating exposure) or affects U.S. reporting via translation exposure. We leave to Appendix B a more complete description of currency exposures as indicated in the firms' financial statements. As shown in the Appendix, firms within an industry display significant common exposures to currency risks. In addition, we note that our firms tend to be large firms (in each industry at most one firm has total assets below the industry median), which avoids the general concern that small firms respond differently to shocks relative to large firms (e.g., Gertler and Gilchrist (1994) for interest rate shocks).

C.1. Oil and Gas Field Machinery and Equipment Manufacturing: NAICS 333132

This industry consists of firms primarily engaged in manufacturing oil and gas field machinery and equipment, such as drilling and production machinery and equipment and oil and gas field derricks, and manufacturing water well drilling machinery. The companies (and dates) included in the analysis are Baker Hughes (1/93 to 12/03), Weatherford International (9/98 to 12/03), Varco International (1/93 to 12/03), CAMCO International (12/93 to 8/98), Cooper Cameron Corp (8/95 to 12/03), National-Oilwell (10/96 to 12/03), IRI International (11/97 to 6/00), NATCO Group (1/00 to 12/03), Grant Prideco (4/00 to 12/03), Oil States International (2/01 to 12/03), and FMC Technologies (6/01 to 12/03). This industry has the most firms and therefore is where we expect to obtain the best results.

This industry is the only one for which we use import weights rather than export weights to construct the currency baskets and to identify the most relevant currencies. We chose import weights because several of the export markets suffered during the sample period due to severe currency crises and a linear model is not a good model for predicting such sporadic events.⁸ In our sample, on average total imports represented 96% of total shipments. The U.S. International Trade Commission database shows that the currencies most relevant in terms of import shares are Canadian dollar, British pound, and all major currencies recently replaced by the euro. To compute the import-weighted exchange rate basket we make use of the following currencies, depending on their relevance: Canadian dollar, British pound, Dutch guilder, French franc, German mark, Italian lira, Austrian schilling, Australian dollar, Argentine peso, Indonesian rupiah, Thai baht, Norwegian krone, and Mexican peso.⁹

⁸ However, this sector's estimated *MPI* does a good job of predicting the Mexican peso and the Indonesian rupiah, but not the Russian ruble.

⁹ Since January 1, 1999, the euro replaced the currencies of the following countries: Belgium, Germany, Spain, France, Ireland, Italy, Luxembourg, the Netherlands, Austria, Portugal, and Finland. Greece joined in January of 2001.

C.2. Aircraft Manufacturing: NAICS 336411

Firms in this industry manufacture or assemble complete aircraft, develop and make aircraft prototypes, convert aircraft (i.e., major modifications to systems), or complete aircraft overhaul and rebuilding (i.e., periodic restoration of aircraft to original design specifications). The companies (and dates) included in the analysis are Boeing (1/93 to 9/03), Gulfstream Aerospace (11/96 to 7/99), Grumman Corp (1/93 to 4/94), and McDonnell Douglas Corp (1/93 to 7/97). Whitehall Corporation is part of this industry but never meets the liquidity criterion. We exclude from the analysis the period 1/00 to 9/03, during which time only Boeing is present.

The dominant currencies associated with this industry are the British pound, Japanese yen, and several European currencies. On aggregate, during our sample period exports were on average 48% of total shipments. To compute the export-weighted exchange rate we make use of the following currencies depending on their relevance in total shipments in each year: British pound, Japanese yen, Australian dollar, Dutch guilder, French franc, German mark, Chinese yuan, Malaysian ringgit, South Korean won, Singapore dollar, Taiwan dollar, and Saudi riyal.

C.3. Aircraft Engine and Engine Parts Manufacturing: NAICS 336412

This industry is engaged in manufacturing aircraft engines and engine parts, in developing and making prototypes of aircraft engines and engine parts, in aircraft propulsion system conversion, and in overhaul and rebuilding. Firms (and dates) included are Heico Corp (1/99 to 12/03), Sequa Corp (1/93 to 12/03), UNC Inc (1/93 to 7/97), United Technologies Corp (1/93 to 12/03), Doncasters PLC (1/97 to 7/01), and Howmet International Inc (11/97 to 6/00), all of which are incorporated in the United States except for Doncasters, which is incorporated in the U.K. We have included Doncasters in our sample because it has a sizeable share of its operations in the U.S.

We calculate the export-weighted exchange rate using the German mark, French franc, British pound, and Canadian dollar. Together these currencies account for over 60% of all exports in our sample, and average exports from this industry represented 50% of total shipments. This aggregate information broadly agrees with the references we found on exchange exposure in company annual reports and 10-K forms, information reported in the Appendix.

C.4. Other Aircraft Parts and Auxiliary Equipment Manufacturing: NAICS 336413

Companies in this industry are primarily engaged in manufacturing aircraft parts or auxiliary equipment (except engines and aircraft fluid power subassemblies) and/or developing and making prototypes of aircraft parts and auxiliary equipment. Auxiliary equipment includes items such as crop dusting apparatus, armament racks, in-flight refueling equipment, and external fuel tanks. The companies (and dates) included in this industry are Honeywell International (1/93 to 12/03), Sundstrand Corp (1/93 to 5/99), Talley Industries (1/93

to 1/98), Triumph Group (10/96 to 12/03), Rockwell Collins Inc. (7/01 to 12/03), Goodrich Corp (1/93 to 12/03), Rohr Industries (1/93 to 12/97), and Ducommun Corp (11/96 to 12/03).

To compute the export-weighted currency return for this industry, we use the following currencies: British pound, Japanese yen, Canadian dollar, French franc, German mark, Korean won, Italian lira, Taiwan dollar, Saudi riyal, and the Israeli shekel. However, the main currencies are the British pound, the Japanese yen, and the Canadian dollar. The ratio of exports to total shipments averaged 57% for the industry.

C.5. Primary Smelting and Refining of Nonferrous Metal (except Copper and Aluminum): NAICS 331419

This industry includes establishments primarily engaged in making nonferrous metals by smelting ore and/or refining nonferrous metals by electrolytic methods or other processes. The four companies (and dates) included in the analysis are Tremont Corp (1/93 to 12/02), WHX (1/93 to 12/02), INCO (1/93 to 12/02), and WMC/Alumina Corp (1/93 to 12/02). We drop 2003 because the large levels of buy and sell orders prevent us from estimating the model. Of the four companies included in the sample, INCO is incorporated in Canada and WMC/Alumina is incorporated in Australia. We include these two foreign incorporated firms because they both have significant operations in the U.S., and INCO, while Canadian, uses the USD as its functional and reporting currency. Unfortunately, Tremont and WMC only rarely fulfill the liquidity criterion in the sample, which means that MPI is almost always estimated with the minimum possible number of firms. As we will see below, this is reflected in the industry's relatively weak results.

The export-weighted exchange rate is constructed using the following currencies: the Swiss franc, British pound, Canadian dollar, Japanese yen, French franc, Taiwan dollar, Hong Kong dollar, Korean won, Australian dollar, and Mexican peso. Of these, the first three currencies are always included in the basket and account for over 70% of all exports from the industry. All other currencies have much smaller weights in terms of trade and are only rarely included in the basket. The ratio of total exports to total shipments is approximately 190%.

III. Results

This section gives details on the maximum likelihood estimation of (5) and the estimation of MPI . It lays out the regression specifications and exact hypothesis tests, and discusses the relevant econometric issues. We show that MPI is not a simple statistical factor of order flow across firms in an industry, and we present the equity and currency returns forecasting results.

A. Maximum Likelihood Estimation of MPI

For each industry, we estimate the value of marketwide private information in month t , MPI_t , by using the vector of estimated parameters

$\hat{\Theta}_t \equiv (\hat{\alpha}_i, \hat{\delta}_i, \hat{\theta}, \hat{\rho}, \hat{\varepsilon}_i^s, \hat{\varepsilon}_i^b, \hat{\mu}_i^a, \hat{\mu}_i^f)_{i=1, \dots, I}$ that maximizes the likelihood function (5) jointly across all firms in month t . This is a constrained maximization with probabilities $0 \leq \alpha_i, \delta_i, \theta, \rho \leq 1$ and trading intensities $0 \leq \hat{\varepsilon}_i^s, \hat{\varepsilon}_i^b, \hat{\mu}_i^a, \hat{\mu}_i^f$. With I firms and 22 trading days per month, we have a total of $I \times 22$ observations to estimate $I \times 6 + 2$ parameters, yielding $I \times 16 - 2$ degrees of freedom. The quality of the estimation therefore increases as the number of firms increases. Significantly, the likelihood maximization uses all the data to estimate each of firm i 's parameters, including data from other firms. The suggestion that only 22 observations from firm i are being used to estimate firm i 's parameters $(\hat{\alpha}_i, \hat{\delta}_i, \hat{\varepsilon}_i^s, \hat{\varepsilon}_i^b, \hat{\mu}_i^a, \hat{\mu}_i^f)$ is incorrect. Intuitively, the estimation of firm i 's parameters is not independent of the estimation of the other parameters because they are linked by the arrival of marketwide private information (see Section I.B for a more detailed argument).

The constrained optimization for which we have to solve is highly dependent on initial conditions, particularly on the parameters μ_i^f and μ_i^a . To minimize the impact of initial conditions, we implement a grid search on the parameters μ_i^f and μ_i^a . Specifically, we set the initial values of these parameters as a fraction (not necessarily the same) of the difference between the maximum daily number of buy orders in a month and the mean daily value of buy orders in that month. The grid allows these fractions to vary between 0.1 and one, with increments of 0.1. We set the initial values of ε_i^s and ε_i^b to each month's mean daily value of sell and buy orders, respectively (Vega (2006)), and set the initial values of the remaining parameters θ, ρ, α_i , and δ_i to 0.5 because we have no prior information regarding these probabilities. The estimated parameters are those that yield the highest value for the likelihood function. For consistency, we apply this procedure over all months and industries. While this procedure is very time-consuming, especially because gradient search methods are of no use here, we are confident that our parameter estimates attain the global maximum. We also tried the EM algorithm that is more robust to initial conditions, but the results do not improve and the estimation slows down considerably.

In order to evaluate the overall fit of the model, we proceed by analyzing the model fit for every estimation, that is, for every month, and also analyze jointly, across months. In the first case, for every month t , the hypothesis we test is

$$H_0^t : \alpha_i = \delta_i = \mu_i^f = \mu_i^a = \theta = \rho = 0, \forall i. \quad (11)$$

Under the null hypothesis, there is no private information and all trading occurs for liquidity reasons. Denote the restricted log-likelihood by $\log L_{0,t}$.¹⁰ The test statistic is a likelihood ratio test, $LR_t = -2(\log L_{0,t} - \log L_t) \sim \chi_{(4 \times I_t + 2)}^2$, where

¹⁰ Under the null, the restricted log-likelihood for month t is

$$\log L = -N \sum_{i=1}^I (\varepsilon_i^s + \varepsilon_i^b) + \sum_{n=1}^N \sum_{i=1}^I (S_{in} \log \varepsilon_i^s + B_{in} \log \varepsilon_i^b) - \sum_{n=1}^N \sum_{i=1}^I \log (S_{in}! B_{in}!).$$

The maximum likelihood estimates in the restricted model are $\varepsilon_i^s = N^{-1} \sum_{n=1}^N S_{in}$, and $\varepsilon_i^b = N^{-1} \sum_{n=1}^N B_{in}$, from which we can construct the maximum log-likelihood, $\log L_{0,t}$.

$\log L_t$ is the log-likelihood of the unrestricted model in month t and I_t is the number of firms in month t .

The second case is a test of the joint hypothesis

$$H_0 : \bigcap_{t=1}^T H_0^t. \tag{12}$$

The joint test H_0 is the intersection of T subhypotheses. Therefore, H_0 is rejected if H_0^t is rejected for at least one t , given the significance levels $\{\alpha_t\}$. However, to test (12) we need the joint distribution of the LR_t statistics from (11), which is unknown unless we restrict the stochastic relationship across subsamples; consequently, the size of the test cannot be determined. To overcome this difficulty, we use the induced test procedure of Dufour and Torrès (1998). According to Dufour and Torrès (1998), we can choose the significance levels α_t of the individual tests so that the induced test has a significance level $\alpha = \sum_{t=1}^T \alpha_t$. A simple rule is to set $\alpha_t = \alpha/T$ for each t . We implement the test as follows. First, we perform the likelihood ratio test for each monthly sample. Second, we choose the month for which we have the lowest p -value on H_0^t and assign the value LR_t of the month's likelihood ratio test statistic to the test statistic of H_0 . Third, the critical value for the month with lowest p -value, $\chi^2_{1-\alpha/T}(4 \times I_t + 2)$, is the critical value of the joint test. Finally, H_0 is rejected if $LR_t > \chi^2_{1-\alpha/T}(4 \times I_t + 2)$.

Table I shows the results of tests (11) and (12). The results show that for almost every month we reject the null of the noise trading model. Also, the joint test strongly rejects the null of no fit of the full model. In summary, these tests present strong evidence in favor of the private information model.

Estimating the model for each month and industry means that we need not worry about possible nonstationarity of information with respect to the

Table I
Quality of Fit of Model Estimation

The table shows the fraction of months in which the noise trading model (hypothesis $H_0^t : \alpha_i = \delta_i = \mu_i^f = \mu_i^a = \theta = \rho = 0, \forall i$) is rejected for each industry. The table also shows the results of the joint significance of the month-by-month estimations. The joint test is the induced test procedure of Dufour and Torrès (1998) on the finite intersection of subhypotheses. The test statistic and critical value for joint hypothesis $H_0 : \bigcap_{t=1}^T H_0^t$ are LR_t and $\chi^2_{1-\alpha/T}(4 \times I_t + 2)$, respectively, which are the value of the likelihood ratio statistic in the month with lowest p -value associated with hypothesis H_0^t and the critical value of the same hypothesis in that month.

	NAICS 333132	NAICS 336411	NAICS 336412	NAICS 336413	NAICS 331419
<i>Individual Goodness of Fit Tests</i>					
Fraction of rejections of H_0^t in %	97.3	98.8	100.0	99.1	100.0
<i>Joint Goodness of Fit Test</i>					
Test statistic	5,216.0	2,241.4	753.6	3,405.6	1572.6
Critical value	74.39	30.96	37.99	57.04	31.90
Number of Observations (Months, T)	114	84	96	121	120

parameters μ_i^f , μ_i^a , ε_i^s , and ε_i^b . However, the observed growth in trading activity over the years requires that we detrend our measures of *MPI*. Easley et al. (2002) endogenously account for this nonstationarity in their estimation. With many firms this process is cumbersome and costly to implement. Hence, we use the Hodrick–Prescott filter for our estimate of order flow driven by marketwide private information. The growth in trading activity and the fact that our firms are generally large firms with a large trading volume imply that estimating *MPI* in the final months of available data is increasingly difficult. Specifically, for some months the number of buy and/or sell orders is so high that maximizing the log-likelihood function requires values higher than the largest positive floating point number in our personal computers. For this reason the sample length varies across industries (see Table I).

B. *MPI* and Factors in Equity Order Flow

Table II characterizes estimated *MPI* by its correlation with other data and model variables. Table II shows that *MPI* is positively correlated with the total industry order flow, *TOF* (i.e., the sum of total buy orders minus the sum of total sell orders over the month). However, this correlation is statistically significant at the 10% level or better for only two of the industries.

We ask whether a simple factor analysis on order flow provides a good measure of marketwide private information that can be used in place of the more complex measure we derive from a structural model (see also Section III.E). Denote by *PC1* the detrended first principal component from intra-industry firm-level order flow in each month. The variable *PC1* accounts for a significant portion of the variability in order flow within an industry. Numbers range from 61.5% in industry 333132 to 95.4% in industry 336412 (untabulated). Not surprisingly, Table II shows that the correlation between *PC1* and *TOF* is almost one. In contrast, we find that the percentage of explained variation in *TOF* from *MPI* is under 10% for all industries (see line 1 in Table II). The results in Table II suggest that *PC1* is not a good proxy for *MPI* because the correlation between *MPI* and *PC1* is low and statistically insignificant in most industries.

We turn next to the question of what explains the large covariation in trades implied by *PC1*. The answer lies in common variation in liquidity trades (ε_i^b and sells ε_i^s). A simple test is the correlation between estimated aggregate liquidity trades $LT_t = \sum_{i=1}^I (\hat{\varepsilon}_{it}^b - \hat{\varepsilon}_{it}^s)$ and *PC1*. These correlations are large, positive, and statistically significant. Further, we construct the first principal component from estimated liquidity trades ($\hat{\varepsilon}_{it}^b - \hat{\varepsilon}_{it}^s$) of the firms in an industry, denoted by *LT1*. The component *LT1* is a better descriptor of common variation in liquidity than is the sum of liquidity trades across firms, *LT*. We ask: How important is *MPI* relative to *LT1* as a source of commonality in trades, as given by *PC1*? We conduct a variance decomposition of the explained variance of *PC1* between *MPI* and *LT1* and assign the covariance term in equal parts to *MPI* and *LT1* because we have no a priori reason to assign it all to either one. We find that *MPI* accounts for 4%, 24%, 25%, and 42% of the explained variation in *PC1* for industries 336411, 336412, 336413, and 331419, respectively, with *LT1*

Table II
Order Flow Correlations

The table shows the contemporaneous correlations between marketwide private information (*MPI*), total order flow (*TOF*), the first principal component of order flow (*PC1*), model estimated aggregate liquidity trades (*LT*), and firm-specific trades (*FPI*) for each industry. All variables are detrended using the Hodrick-Prescott filter. The second value in each cell is the associated autocorrelation- and heteroskedasticity-adjusted *p*-value for the null that the correlation coefficient is zero. An asterisk (*) indicates significance at 10% or better.

Correlation Between	NAICS 333132	NAICS 336411	NAICS 336412	NAICS 336413	NAICS 331419
MPI and TOF	0.3258*	0.1602	0.1578	0.1783	0.3180*
	0.0785	0.2785	0.3765	0.1279	0.0120
MPI and PC1	0.2911*	0.0447	0.1299	0.1244	0.3016*
	0.0817	0.7791	0.3936	0.1950	0.0200
MPI and LT	-0.6585*	-0.6678*	-0.7173*	-0.5281*	0.0917
	0.0000	0.0000	0.0000	0.0002	0.4498
MPI + LT and TOF	0.6401*	0.7209*	0.5002*	0.6612*	0.3517*
	0.0000	0.0000	0.0000	0.0000	0.0036
MPI + LT and PC1	0.4889*	0.7220*	0.5197*	0.6265*	0.3853*
	0.0001	0.0000	0.0000	0.0000	0.0012
LT and TOF	0.3426*	0.5039*	0.2805*	0.5170*	0.2643*
	0.0272	0.0007	0.0340	0.0000	0.0343
LT and PC1	0.2418*	0.5516*	0.3091*	0.3581*	0.3080*
	0.0958	0.0001	0.0147	0.0321	0.0103
FPI and TOF	0.4214*	0.2408	0.0678	0.1956*	0.2259*
	0.0052	0.1755	0.6311	0.0899	0.0589
FPI and MPI	0.6874*	0.6645*	0.7373*	0.5133*	0.2400*
	0.0000	0.0000	0.0000	0.0002	0.0833
FPI and LT	-0.6334*	-0.6344*	-0.8935*	-0.6861*	0.0676
	0.0000	0.0000	0.0000	0.0007	0.5491
FPI and PC1	0.4732*	0.1511	0.0363	0.1770*	0.2287*
	0.0038	0.4533	0.7985	0.0393	0.0517
MPI+FPI+LT and TOF	0.9767*	0.9688*	0.7437*	0.9824*	0.3751*
	0.0000	0.0000	0.0000	0.0000	0.0042
MPI+FPI+LT and PC1	0.8708*	0.9024*	0.7287*	0.9364*	0.4080*
	0.0000	0.0000	0.0000	0.0000	0.0015
PC1 and TOF	0.8736*	0.8757*	0.9752*	0.9474*	0.9632*
	0.0000	0.0000	0.0000	0.0000	0.0000
Number of Observations	114	84	96	121	120

accounting for the remaining fraction (untabulated). The exception is industry 333132, where *MPI* explains 98% of the variation in *PC1*, whereas *LT1* explains only 2%. Therefore, the results indicate that in four out of five industries common variation in liquidity trades is the most important source of common variation in order flow. This finding can shed light on the statistical exercise in Hasbrouck and Seppi (2001) that documents significant co-movement in order flows of the 30 stocks in the Dow Jones Industrial.

Table II shows that *MPI* (and also *FPI*) is negatively correlated with *LT*. According to our model this correlation should be zero because the arrival of

uninformed trades is independent of the arrival of informed trades. The negative correlation arises because of small sample estimation error combined with the fact that the estimation captures total order flow quite well. (Appendix C presents a formal analysis of this result.) Intuitively, as the model tries to match the total number of buy and sell orders, more orders explained by trading due to private information lead to less being attributed to liquidity trading. To illustrate this intuition, suppose, for simplicity, that no trades due to firm-specific private information exist and that all marketwide private information is good news. For simplicity also suppose that θ is known. As a result, the maximum likelihood estimate of ε_i^s is $N^{-1} \sum_{n=1}^N S_{in}$ because nothing else explains why investors sell. In contrast, the buy orders can either come from μ_i^a or ε_i^b . As the model tries to capture the total number of buy orders $N^{-1} \sum_{n=1}^N B_{in}$, any estimation error in μ_i^a affects negatively estimates of ε_i^b , but not estimates of ε_i^s . Hence, in this case, estimates of $LT_i = \varepsilon_i^b - \varepsilon_i^s$ vary negatively with estimates of informed trading $MPI_i = \theta \mu_i^a$.

We do not think that estimation error, in either μ_i^a , ε_i^b or ε_i^s , matters for the interpretation of other correlations emphasized in this paper. First, we expect that estimation error will not affect the interpretation of the correlation between estimated MPI or LT and the first principal component in order flow, $PC1$. This is because estimates of $PC1$ are obtained outside of the maximum likelihood estimation. However, because overestimation of MPI is more likely when the likelihood identifies too much correlation in trades, that is, in months when $PC1$ is also high, if anything we expect estimation error to increase the correlation between MPI and $PC1$. We find that the correlation between estimated LT and $PC1$ is higher than that between MPI and $PC1$, and consequently our results are robust to measurement error. Second, we expect any small sample bias in estimating MPI to go against finding evidence of any forecasting power.

The sum of MPI with LT is highly correlated with total order flow TOF , but $MPI + LT$ is not the model's average total trading. For that we also need the model's estimate of trading due to firm-specific private information, $FPI_t = \sum_{i=1}^I \hat{\alpha}_{i,t}(1 - 2\hat{\delta}_{i,t})\hat{\mu}_{i,t}^f$. The model captures total trading activity quite well: The sum $MPI + FPI + LT$ is highly correlated with TOF (correlations around 0.97 for three of the industries, 0.74 for 336412 and 0.37 for 331419) and also with $PC1$. The high correlation between $MPI + FPI + LT$ and TOF gives us confidence that, at least for the first four industries, the model appears to decompose order flow well and confirms in a different way the goodness of fit results described in the previous section.

The last column of Table II gives the numbers for industry 331419. Looking across industries makes apparent that this industry's MPI stands out and contrasts with the MPI for the other industries. We believe this is related to this industry's use of two firms most of the time to estimate MPI , with the other two firms in the industry sporadically meeting the liquidity criterion (see Section II.C).

Harford and Kaul (2005) present evidence that common effects in order flow measured from factor analysis are strong across stocks that belong to the same

index, but are economically inconsequential in nonindex stocks. Industry effects exist but are also small. Even though our data show *MPI* as weakly correlated or even uncorrelated with the first principal component, Harford and Kaul's evidence might still complicate our analysis if our industry composition relied heavily on index constituent firms. However, this is generally not the case in our data: 7 out of the 11 companies in NAICS 333132 are index constituents, only 1 out of 6 firms in NAICS 336411 belongs to an index, only 1 out of 6 in NAICS 336412 belongs to an index, 4 out of 8 firms in NAICS 336413 belong to an index, and none of the 4 companies in NAICS 331419 is an index constituent.¹¹

Finally, we ask how estimated *MPI* is correlated across industries. The answer to this question is important to help make the case that *MPI* identifies marketwide private information as opposed to industry-wide private information. In the latter case we expect the first principal component across all *MPI*s to explain 20% (i.e., one over the number of industry factors) of the variation in *MPI*s. Instead, a principal component analysis on the five estimated values of *MPI* reveals that the first component explains 46% of the total variation and that the first two components explain 81% of the total variation. Iterated principal factor analysis reveals that the first factor explains 58% of the total variation in *MPI*s and the first two factors explain 80%. This seems to indicate that at most two sources of marketwide private information affect these five industries. However, the absence of more economic structure in the model makes further evaluation of this hypothesis difficult; even if only one factor is present different industry sensitivities to the factor and noise in estimated *MPI* can make the correlation across *MPI*s low.

C. Specification of the Forecasting Regressions and Hypothesis Tests

C.1. Forecasting Equity Returns

For each industry, we regress firm stock returns on lagged values of *MPI*:

$$RET_{i,t+j} = a_{i,0} + \sum_{l=1,\dots,L} a_l MPI_{t-l} + u_{i,t+j}. \quad (13)$$

In each regression $RET_{i,t+j}$ is either the j -month-ahead holding period stock return with $j = 1, 2$, or the 60-, 90-, or 120-day ahead cumulative return for firm i . The lag length L is determined via the Akaike Information Criterion (AIC) or by the Bayesian Information Criterion (BIC) whenever the former is ambiguous.¹² The inclusion of lags in this regression follows Hasbrouck (1991)

¹¹ Indexes considered are S&P 500, S&P MidCap 400, and S&P SmallCap 600. Mostly, the firms in our sample that were index constituents belonged to the S&P 500.

¹² By this we mean that whenever the AIC turned out to be of the same magnitude for two or more lag lengths. Choosing in accordance with the BIC means, as to be expected, the choice of the more parsimonious specification. This also means that using the BIC instead of the AIC in every situation would deliver specifications at odds with the Wald significance tests.

and is meant to capture permanent information effects as opposed to temporary inventory effects. The firms whose returns we attempt to forecast are the same as those we use to estimate *MPI*. Hypothesis 1 is that *MPI* forecasts equity returns,

$$H_0 : \alpha_1 = \dots = \alpha_L = 0, \quad (14)$$

against $H_A : \alpha_l \neq 0$, for all $l > 0$. Because a positive value of *MPI* is indicative of good news for firms in the industry, we also look for a positive cumulative response of equity returns to shocks to *MPI*: $\sum_{l=1}^L \alpha_l > 0$.

As indicated in (13), we use panel data methods to test the impact of *MPI* on equity returns. Specifically, we assume fixed effects in the intercept and use the Within-Groups estimator. The generalized Hausman test procedure rejects both the random effects specification and the random coefficients specification in every case. The alternative is to aggregate firm returns into an industry return (based on the firms we use), but the model does not give any guidance on how to do this (e.g., value weights, export-share weights, or simple weights). On the other hand, pooling the data has the advantages of guaranteeing that *MPI* is forecasting returns (not the industry weights) and of rendering more degrees of freedom.

C.2. Forecasting Currency Returns

For each industry, we use either currency returns or excess currency returns in the regression specifications. Let ΔCUR_{t+j} be either the value of the currency return at time $t+j$ or the value of the excess currency return at time $t+j$. Our regression model is

$$\Delta CUR_{t+j} = \alpha_0 + \sum_{l=1, \dots, 10} \alpha_l MPI_{t-l+1} + u_{t+j}. \quad (15)$$

The regressions are conducted with contemporaneous ($j = 0$), 1-month ahead ($j = 1$), and 2-months ahead ($j = 2$) currency and excess currency returns. The hypothesis we test (Hypothesis 2) is that *MPI* explains currency movements,

$$H_0 : \alpha_1 = \dots = \alpha_{10} = 0, \quad (16)$$

against $H_A : \alpha_l \neq 0$, for all $l > 0$.

In specification (15), we add 9 months of lags of *MPI* beyond the contemporaneous value. This lag length was optimal in the AIC sense for most regressions, but not all. We are aware this might mean that some regressions are overfitted with an associated loss of power. However, fixing the number of lags across forecasting horizons and currencies facilitates the interpretation of the regression R^2 s. Furthermore, the AIC rule is often at odds with the Wald significance test, leading us to reject globally significant regressions in favor of more parsimonious but nonsignificant specifications.

We also estimate a constrained version of (15), where $\alpha_2 = \dots = \alpha_{10} = 0$. This specification presumes that private information diffuses quickly and is fully

incorporated into prices in one month. Instead, the specification in (15) assumes that new information about future fundamentals summarized by the component of order flow associated with marketwide private information is only gradually impounded into prices and learned by investors. While this constrained specification generates substantial predictability, we find that the data strongly favor the gradual diffusion of information hypothesis, and we report the latter results alone to conserve space.

The use of a generated regressor in (13) and (15) means that we have to account for possible errors in variables. As shown in Pagan (1984) and Newey and McFadden (1994), least squares in regressions with generated regressors remains consistent in most cases if the nuisance parameters have been consistently estimated in a previous step. However, the sampling variation of the estimator used to generate the regressor may affect inference, which would require the correction of the standard errors for the first-step estimation (Newey and McFadden (1994)).¹³ To this end, the usual heteroskedasticity- and autocorrelation-robust estimate of the covariance matrix can be used if one assumes weak exogeneity of the disturbance term in the regression with respect to the data used to estimate our measure of private information so long as $\alpha_l = 0$. The assumption of weak exogeneity is justified in the regression equation (15) because the explanatory variable is a measure of trade volume in the stock market, whereas our dependent variable is a return in the foreign exchange market. Under the null $H_0 : \alpha_l = 0, l = 1, \dots, L$, the uncorrected Newey-West estimator of the asymptotic variance of the OLS estimator of regression coefficients in (15) is consistent (likewise for model (13) under the null $H_0 : \alpha_l = 0, l = 1, \dots, L$). Note that we use robust standard errors because of the autocorrelation induced by overlapping observations (see Hansen and Hodrick (1980)).

D. Forecasting Results

We start with the results for equity returns in Table III. Each cell of Table III has three numbers: the first is the sum of the estimated coefficients $\sum_{l=1}^L \hat{\alpha}_l$, the second is the p -value on the significance of the sum of the coefficients, and the third is the p -value on the null hypothesis $H_0 : \alpha_l = 0, l = 1, \dots, L$. Our estimated information shock, MPI , successfully forecasts returns in four out of five industries. For industries 333132, 336411, 336412, and 336413, not only do we reject the null of no explanatory power of MPI , but we also obtain positive, significant estimates for the cumulative response of equity returns.¹⁴ For

¹³ The use of bootstrapping is an alternative. In this case, however, this is not feasible because it requires bootstrapping not only the second-step estimation, but also simultaneously the first and time-consuming MLE step.

¹⁴ Stambaugh (1999) shows that the significance of regressions with asset returns as dependent variables can be overstated. This overstatement increases if one regresses stock returns on a lagged stochastic regressor that depends on prices, such as the dividend yield, because the OLS estimator will have an upward finite-sample bias. As Stambaugh argues, by definition such an explanatory variable will not be orthogonal to the disturbance term in the predictive regression at all leads and

Table III
Stock Returns and Order Flow Driven by Marketwide
Private Information

The table shows statistics obtained from the regression

$$RET_{i,t+j} = a_{i,0} + \sum_{l=1,\dots,L} a_l MPI_{t-l+1} + u_{i,t+j}.$$

In Panel A, $RET_{i,t+j}$ is the monthly return in month $t+j$ for stock i , and in Panel B $RET_{i,t+j}$ is the 60-day, 90-day, and 120-day cumulative return for the same stock, in month t . MPI_t is the estimated measure of order flow driven by marketwide private information in month t . The model is estimated using panel data and the Within-Groups estimator. The first number in each cell reports the estimate of $\sum_{l=1,\dots,L} a_l$, the second number gives the p -value on the hypothesis that the sum of the coefficients is zero, and the third number reports the p -value of a Wald test on the null hypothesis of joint significance, $H_0 : a_l = 0, l > 0$. L is chosen using Akaike's Information Criterion or, whenever ambiguous, the Bayesian Information Criterion. p -values are obtained with robust standard errors. An asterisk (*) indicates significance at 10% or better.

<i>j</i>	NAICS 333132	NAICS 336411	NAICS 336412	NAICS 336413	NAICS 331419
Panel A: Single-Period Returns					
1	0.0007*	0.0004	−0.0000	0.0027*	0.0008
	0.0140	0.6660	0.9900	0.0620	0.3754
	0.0000*	0.0000*	0.0024*	0.0004*	0.3754
2	0.0007*	0.0003*	0.0020*	0.0043*	−0.0005
	0.0080	0.0718	0.0110	0.0060	0.5531
	0.0000*	0.0718*	0.0693*	0.0002*	0.5531
Panel B: Cumulative Returns					
60-day	0.0014*	0.0006*	0.0269*	0.0073*	0.0002
	0.0000	0.0320	0.0020	0.0010	0.8629
	0.0000*	0.0672*	0.0934*	0.0000*	0.8629
90-day	0.0022*	0.0010*	0.0030*	0.0110*	0.0007
	0.0000	0.0030	0.0040	0.0000	0.5478
	0.0000*	0.0100*	0.0118*	0.0000*	0.5478
120-day	0.0030*	0.0008*	0.0023*	0.0152*	0.0013
	0.0000	0.0024	0.0060	0.0000	0.3334
	0.0000*	0.0024*	0.0150*	0.0000*	0.3334

1-month ahead stock return regressions in industries NAICS 336411 and 336412 we find globally significant regressions, but with a statistically insignificant sum of coefficients associated with current and lagged MPI . On the negative side, we cannot forecast returns in industry 331419 with MPI . While we are not certain of the reason for this failure, we note our discussion in Section III.B on the reasons this industry may have trouble generating a good estimate of

lags. We do not feel the need to correct for this potential bias because our explanatory variable, MPI , is a measure of trade volume that does not directly depend on stock prices. This is even more true for the currency return regressions because MPI relates to a different asset market. Furthermore, there is controversy on the use of this correction (Lewellen (2004)).

marketwide private information. We also note that this industry displays the highest volatility of stock returns with a coefficient of variation that is almost more than double that of the others (13.09 vs. 4.1 to 7.2 for the other industries) and displays the highest kurtosis (11.25 vs. 4.8 to 7 for the other industries), making it harder for *MPI* to explain returns.

It is instructive to compare the speed of information diffusion across measured marketwide and firm-specific information events. Consider identifying information events by the absolute value of *FPI* for firm-specific information and by the absolute value of *MPI* for marketwide information. We determine the rate of information diffusion by running a time-series regression of the absolute value of firm *i*'s equity return on $|FPI_i|$, or $|MPI_i|$ and their lagged values, respectively (Easley and O'Hara (1992) show convergence in a similar setup). To maximize the number of observations, the analysis is conducted using only firms with complete presence in the sample. The results (untabulated) show that the model is estimated well for all but one firm and are consistent with the prediction that firm-specific private information diffuses faster than marketwide private information. The reason might be that firm-specific information is generally more objective, easier to interpret, and easier to trade on than marketwide private information.

Turn to the forecasting of currency returns in Tables IV to VIII.¹⁵ The summary of our results is that *MPI* forecasts currency returns and excess currency returns quite well: (i) *MPI* forecasts excess currency-basket returns slightly better than simple currency-basket returns; (ii) *MPI* can forecast currency-basket returns but also, importantly, returns of the main individual currencies that compose these baskets; (iii) lags of *MPI* appear to have significant forecasting ability consistent with slow diffusion of information over time and also consistent with Evans and Lyons (2004b); and (iv) the R^2 s in these regressions are in the range of 5% to 25%.

Panels A and B of Table IV present results for the Oil and Gas Field Machinery and Equipment Manufacturing industry (NAICS 333132). The first figure in each cell reports the R^2 of the regression, and the second the *p*-value associated with the hypothesis test as described in Section III.C above. (This structure is repeated in Tables IV to VIII.) Because of the large number of firms with apparent similar exposures in this industry, we expect that this industry is one such that *MPI* is estimated well. Indeed, we find that the R^2 s in the forecasting regressions in this industry are quite large. Contemporaneous and lagged *MPI* have predictive power for the contemporaneous currency returns; 1-month- and two-month-ahead currency returns; and excess returns on every currency, including the currency basket. The exception is the British pound. Note that

¹⁵ Recall that the individual currencies we study were chosen if they had a large weight in the trade-weighted currency basket consistently over the sample period, or if they were referred by the firms as a source of significant foreign exchange exposure. Some currencies like the Argentine peso, Brazilian real, and Indonesian rupiah, though sometimes referenced by companies, were excluded because of extreme devaluation episodes that invalidate inference with linear models. They were nonetheless included in baskets if our criterion called for them, and if they were outside these devaluation periods.

Table IV
Exchange Rates and Order Flow Driven by Marketwide Private Information in the Oil and Gas Field Machinery and Equipment Manufacturing Industry, NAICS 333132

The table shows statistics obtained from the regression

$$\Delta CUR_{t+j} = \alpha_0 + \sum_{l=1,\dots,10} \alpha_l MPI_{t-l+1} + u_{t+j}.$$

In Panel A ΔCUR_{t+j} is the currency return in month $t+j$, in Panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t . The first number in each cell reports the R^2 of the regression, and the second number reports the autocorrelation- and heteroskedasticity-adjusted p -value of a Wald test on the null hypothesis of joint significance, $H_0 : \alpha_l = 0, l > 0$. The sample is January 1993 to June 2002. The first column reports results when CUR is a trade-weighted exchange rate (Basket). The currencies in the basket are indicated in the main text and the weights correspond to the import share into each of these countries adjusted to add to one. Average import shares are shown under the currency name. In the remaining columns CUR is one of the following currencies: Canadian Dollar (CAD), German Mark (DEM), French Franc (FRF), British Pound (GBP), Italian Lira (ITL), Dutch Guilder (NLG), and the Norwegian Kroner (NOK). An asterisk (*) indicates significance at 10% or better.

	Currency Basket	CAD 26.09%	DEM 5.33%	FRF 6.87%	GBP 17.63%	ITL 5.74%	NLG 7.79%	NOK 3.94%
Panel A: Currency Returns								
$j = 0$	0.1200*	0.1272*	0.1417*	0.1451*	0.0575	0.1822*	0.1445*	0.1370*
	0.0002	0.0000	0.0000	0.0000	0.6866	0.0000	0.0000	0.0000
$j = 1$	0.1644*	0.1796*	0.1585*	0.1653*	0.0452	0.1928*	0.1620*	0.1433*
	0.0000	0.0000	0.0000	0.0000	0.7097	0.0000	0.0000	0.0000
$j = 2$	0.1184*	0.1621*	0.1371*	0.1397*	0.0490	0.1732*	0.1394*	0.1463*
	0.0015	0.0027	0.0002	0.0001	0.1059	0.0000	0.0001	0.0001
Panel B: Excess Currency Returns								
$j = 0$	0.1107*	0.1259*	0.1397*	0.1433*	0.0584	0.1821*	0.1428*	0.1309*
	0.0009	0.0000	0.0000	0.0000	0.6313	0.0000	0.0000	0.0000
$j = 1$	0.1672*	0.1800*	0.1607*	0.1674*	0.0450	0.1952*	0.1649*	0.1384*
	0.0005	0.0000	0.0000	0.0000	0.6704	0.0000	0.0000	0.0000
$j = 2$	0.1288*	0.1641*	0.1419*	0.1442*	0.0497*	0.1790*	0.1252*	0.1447*
	0.0005	0.0018	0.0000	0.0000	0.0930	0.0000	0.0000	0.0000

MPI forecasts returns on the Canadian dollar, the most important currency in the industry.

Table V gives the results for the Aircraft Manufacturing industry (NAICS 336411). Here MPI shows significant predictive power for most currencies and specifications. Across all selected currencies, only for contemporaneous and 1-month-ahead returns of the British pound and for simple returns of the currency basket do we not find any significant correlation with MPI . In particular, for the Canadian dollar and the Japanese yen we find R^2 s in excess of 20%, and in most instances the regressions are globally significant at the 1% level of significance.

Table V

Exchange Rates and Order Flow Driven by Marketwide Private Information in the Aircraft Manufacturing Industry, NAICS 336411

The table shows statistics obtained from the regression

$$\Delta CUR_{t+j} = \alpha_0 + \sum_{l=1,\dots,10} \alpha_l MPI_{t-l+1} + u_{t+j}.$$

In Panel A ΔCUR_{t+j} is the currency return in month $t+j$, in Panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t . The first number in each cell reports the R^2 of the regression, and the second number reports the autocorrelation- and heteroskedasticity-adjusted p -value of a Wald test on the null hypothesis of joint significance, $H_0 : \alpha_l = 0, l > 0$. The sample is January 1993 to December 1999. The first column reports results when CUR is a trade-weighted exchange rate (Basket). The currencies in the basket are indicated in the main text and the weights correspond to the export share into each of these countries adjusted to add to one. Average export shares are shown under the currency name. In the remaining columns CUR is one of the following currencies: Canadian Dollar (CAD), German Mark (DEM), French Franc (FRF), British Pound (GBP), Japanese Yen (JPY), and the Dutch Guilder (NLG). An asterisk (*) indicates significance at 10% or better.

	Currency Basket	CAD 2.16%	DEM 4.57%	FRF 2.57%	GBP 9.31%	JPY 8.23%	NLG 3.79%
Panel A: Currency Returns							
$j = 0$	0.0908	0.2222*	0.1157*	0.1117*	0.0505	0.2535*	0.1184*
	0.4332	0.0000	0.0365	0.0505	0.4616	0.0000	0.0357
$j = 1$	0.1198	0.2059*	0.1413*	0.1493*	0.0605	0.2682*	0.1449*
	0.4562	0.0000	0.0001	0.0010	0.3705	0.0000	0.0001
$j = 2$	0.1207	0.2086*	0.1486*	0.1501*	0.1790*	0.2138*	0.1427*
	0.6177	0.0000	0.0000	0.0001	0.0461	0.0000	0.0000
Panel B: Excess Currency Returns							
$j = 0$	0.1059*	0.2350*	0.1149*	0.1084*	0.0560	0.2551*	0.1192*
	0.0200	0.0000	0.0415	0.0847	0.3319	0.0000	0.0346
$j = 1$	0.1103*	0.2180*	0.1414*	0.1464*	0.0630	0.2703*	0.1464*
	0.0197	0.0000	0.0000	0.0007	0.2948	0.0000	0.0001
$j = 2$	0.0735*	0.2153*	0.1499*	0.1502*	0.1837*	0.2150*	0.1455*
	0.0072	0.0000	0.0000	0.0001	0.0321	0.0000	0.0000

Table VI displays the results for the Aircraft Engine Manufacturing industry (NAICS 336412). In this industry, MPI does a poor job forecasting Canadian dollar returns, but it does particularly well for the currency basket and the most important currencies in the industry, the British pound and the Japanese yen. We can still forecast contemporaneous and 1-month-ahead currency returns for the French franc, but only contemporaneous ones for the German mark. Interestingly, contemporaneous MPI alone can forecast 1-month-ahead returns for the German mark, that is, for $j = 1$ (results available upon request). The failure of the distributed lag structure to explain 1-month-ahead returns for the German mark is the exception in our results and could be due to either overfitting of the regression or to the loss of observations due to the inclusion of the lags.

Table VI
Exchange Rates and Order Flow Driven by Marketwide Private Information in the Aircraft Engine and Engine Parts Manufacturing Industry, NAICS 336412

The table shows statistics obtained from the regression

$$\Delta CUR_{t+j} = \alpha_0 + \sum_{l=1,\dots,10} \alpha_l MPI_{t-l+1} + u_{t+j}.$$

In Panel A ΔCUR_{t+j} is the currency return in month $t+j$, in Panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t . The first number in each cell reports the R^2 of the regression, and the second number reports the autocorrelation- and heteroskedasticity-adjusted p -value of a Wald test on the null hypothesis of joint significance, $H_0 : \alpha_l = 0, l > 0$. The sample is January 1993 to December 2000. The first column reports results when CUR is a trade-weighted exchange rate (Basket). The currencies in the basket are indicated in the main text and the weights correspond to the export share into each of these countries adjusted to add to one. Average export shares are shown under the currency name. In the remaining columns CUR is one of the following currencies: Canadian Dollar (CAD), German Mark (DEM), French Franc (FRF), British Pound (GBP), and Japanese Yen (JPY). An asterisk (*) indicates significance at 10% or better.

	Currency Basket	CAD 11.84%	DEM 8.44%	FRF 22.89%	GBP 14.24%	JPY 6.13%
Panel A: Currency Returns						
$j = 0$	0.1482*	0.0447	0.1267*	0.1337*	0.1095*	0.1138*
	0.0010	0.8716	0.0137	0.0181	0.0517	0.0012
$j = 1$	0.1227*	0.0482	0.1114	0.1186*	0.0999*	0.1064*
	0.0174	0.7887	0.1031	0.0663	0.0043	0.0023
$j = 2$	0.1335*	0.0932	0.1203	0.1259	0.1293*	0.0684*
	0.0188	0.5109	0.2187	0.2032	0.0000	0.0459
Panel B: Excess Currency Returns						
$j = 0$	0.1528*	0.0444	0.1276*	0.1342*	0.1107*	0.1137*
	0.0006	0.8749	0.0183	0.0217	0.0524	0.0009
$j = 1$	0.1256*	0.0504	0.1119	0.1184*	0.0998*	0.1065*
	0.0045	0.7891	0.1431	0.0769	0.0053	0.0022
$j = 2$	0.1422*	0.0933	0.1230	0.1278	0.1307*	0.0681*
	0.0071	0.5189	0.2154	0.2188	0.0000	0.0372

Table VII shows the results for the Aircraft Parts Manufacturing industry (NAICS 336413). Here MPI forecasts currency returns and excess currency returns for every currency at several horizons, except for the Japanese yen. It can also forecast simple currency returns 1 and 2 months ahead for the currency basket. However, when compared with the three previous industries, MPI accounts for a somewhat smaller share of each currency return's total variation: overall R^2 s are somewhat lower. We fail to find predictability for $j = 1$ for the DEM and FRF simple currency returns, but because predictability in excess currency returns is not rejected for $j = 1$, we do not think that rejection of predictability for simple returns is due to misidentification of currency exposures. Instead, the impact of measurement error may be more strongly felt (note that

Table VII

Exchange Rates and Order Flow Driven by Marketwide Private Information in the Other Aircraft Parts and Auxiliary Equipment Manufacturing Industry, NAICS 336413

The table shows statistics obtained from the regression

$$\Delta CUR_{t+j} = \alpha_0 + \sum_{l=1,\dots,10} \alpha_l MPI_{t-l+1} + u_{t+j}.$$

In Panel A ΔCUR_{t+j} is the currency return in month $t+j$, in Panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t . The first number in each cell reports the R^2 of the regression, and the second number reports the autocorrelation- and heteroskedasticity-adjusted p -value of a Wald test on the null hypothesis of joint significance, $H_0 : \alpha_l = 0, l > 0$. The sample is January 1993 to February 2003, excluding the month of October 2000. The first column reports results when CUR is a trade-weighted exchange rate (Basket). The currencies in the basket are indicated in the main text and the weights correspond to the export share into each of these countries adjusted to add to one. Average export shares are shown under the currency name. In the remaining columns CUR is one of the following currencies: Canadian Dollar (CAD), German Mark (DEM), French Franc (FRF), British Pound (GBP), and Japanese Yen (JPY). An asterisk (*) indicates significance at 10% or better.

	Currency Basket	CAD 9.38%	DEM 5.31%	FRF 6.38%	GBP 14.24%	JPY 14.01%
Panel A: Currency Returns						
$j = 0$	0.0495	0.1045	0.0651*	0.0694*	0.1175*	0.0329
	0.6534	0.4767	0.0617	0.0121	0.0017	0.8911
$j = 1$	0.0781*	0.0941*	0.0606	0.0632	0.1033*	0.0258
	0.0076	0.0648	0.2071	0.1103	0.0006	0.8800
$j = 2$	0.0634*	0.0773*	0.0759*	0.0703*	0.1143*	0.0557
	0.0084	0.0706	0.0006	0.0042	0.0000	0.6035
Panel B: Excess Currency Returns						
$j = 0$	0.0522	0.1054	0.0783*	0.0727*	0.1171*	0.0342
	0.6199	0.4795	0.0047	0.0108	0.0021	0.8920
$j = 1$	0.0520	0.0960*	0.0717*	0.0676*	0.1042*	0.0264
	0.6556	0.0549	0.0492	0.0827	0.0004	0.8718
$j = 2$	0.0774*	0.0775*	0.0792*	0.0740*	0.1141*	0.0566
	0.0100	0.0649	0.0003	0.0018	0.0000	0.6170

predictability in simple FRF returns is only marginally rejected) than in most cases.

Finally, Table VIII presents the results for the Primary Smelting of Nonferrous Metal industry (NAICS 331419). Contemporaneous and lagged MPI forecast the currency basket in terms of simple and excess returns at almost every horizon, in spite of the difficulty in forecasting stock returns in this industry. In terms of individual currencies, we do well estimating the Swiss franc—the single most important currency in terms of exports—the French franc, and the Japanese yen. Conversely, we fail to forecast the second-most important currency, the British pound, and, again, the R^2 s are considerably lower than in the

Table VIII
Exchange Rates and Order Flow Driven by Marketwide Private Information in the Primary Smelting and Refining of Nonferrous Metal, NAICS 331419

The table shows statistics obtained from the regression

$$\Delta CUR_{t+j} = \alpha_0 + \sum_{l=1,\dots,10} \alpha_l MPI_{t-l+1} + u_{t+j}.$$

In Panel A ΔCUR_{t+j} is the currency return in month $t+j$, in Panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t . The first number in each cell reports the R^2 of the regression, and the second number reports the autocorrelation- and heteroskedasticity-adjusted p -value of a Wald test on the null hypothesis of joint significance, $H_0 : \alpha_l = 0, l > 0$. The sample is January 1993 to December 2002. The first column reports results when CUR is a trade-weighted exchange rate (Basket). The currencies in the basket are indicated in the main text and the weights correspond to the export share into each of these countries adjusted to add to one. Average export shares are shown under the currency name. In the remaining columns CUR is one of the following currencies: Canadian Dollar (CAD), Swiss Franc (CHF), French Franc (FRF), British Pound (GBP), and the Japanese Yen (JPY). An asterisk (*) indicates significance at 10% or better.

	Currency Basket	CAD 6.18%	CHF 38.21%	FRF 2.62%	GBP 26.59%	JPY 5.10%
Panel A: Currency Returns						
$j = 0$	0.0879*	0.0872	0.0614*	0.0853*	0.0435	0.0588*
	0.0212	0.4992	0.0740	0.0383	0.3981	0.0060
$j = 1$	0.0854*	0.0933	0.0674*	0.0896*	0.0429	0.0579*
	0.0156	0.1950	0.0601	0.0173	0.4400	0.0040
$j = 2$	0.0698*	0.0858	0.0458	0.0383	0.0417	0.0604*
	0.0761	0.1230	0.5566	0.6835	0.2890	0.0031
Panel B: Excess Currency Returns						
$j = 0$	0.0881*	0.0881	0.0631*	0.0872*	0.0428	0.0565*
	0.0094	0.4571	0.0513	0.0251	0.7840	0.0053
$j = 1$	0.0877*	0.0927	0.0710*	0.0936*	0.0439	0.0574*
	0.0076	0.1593	0.0404	0.0094	0.4311	0.0030
$j = 2$	0.0777*	0.0861*	0.0471	0.0389	0.0417	0.0600*
	0.0282	0.0975	0.5083	0.6434	0.2920	0.0023

other industries. The poor performance of this industry is consistent with our suspicion that the model cannot estimate MPI properly.

Our results reveal that the speed of information diffusion is not necessarily the same across industries and across currencies. This could be due to a variety of reasons. First, a faster speed of information diffusion could arise because marketwide information in the industry/currency is in general more accessible to traders (Vives (1993)). Second, the marketwide information contained in MPI possibly is revealed through public signals more quickly in some industries. Third, a faster speed of information diffusion arises if the currency order flow that follows such marketwide events contains fewer liquidity trades, thus impacting prices faster and generating less hot potato trading (Lyons (1997)).

E. Robustness Checks

To check the robustness of the results presented in the previous sections we conduct several additional regressions and tests.

Table II documents a small but positive correlation of *MPI* with total industry order flow. With very few exceptions, current total order flow shows no predictive power in explaining current or future (up to 2-months-ahead) currency returns or excess currency returns at the 10% significance level. These results imply that our measure of marketwide private information can better extract private information than the simple averaging out of order flow.

Another robustness check is the running of a horse race between *MPI* and the first principal component of total order flow, *PC1*. Recall from our discussion above that this statistical factor captures all comovement in order flow across firms, including some derived from marketwide private information. In this robustness exercise we determine which measure does a better job of forecasting currency returns by analyzing if the inclusion of *PC1* negatively affects *MPI*'s forecasting ability. We regress each currency return on contemporaneous and lagged *MPI* and on contemporaneous and lagged *PC1*. For a neutral comparison we use nine lags for both *MPI* and *PC1*.

Untabulated results (available upon request) show that *MPI* does not lose explanatory power with the introduction of *PC1*. In every industry, contemporaneous and lagged *MPI* remain jointly significant, in cases where this was previously the case, whereas *PC1* only does so for industries 336413 and 331419 and to a lesser extent for industry 333132. Recall that industries 336413 and 331419 are those for which *MPI* did not do as well. When we look at the share of the dependent variable's total variation explained by *MPI*, we note that it decreases at most 12% (less than one percentage point) on average for industry 331419 and about 8% (or 1.2 percentage points) for industry 333132. For the remaining industries we observe an increase in *MPI*'s explanatory power after the introduction of contemporaneous and lagged *PC1* in the list of explanatory variables. We conclude that *PC1* is capturing some other source of explanatory power because *MPI*'s forecasting ability does not seem to be significantly affected by the principal component.

What are the potential sources of the explanatory power of *PC1* for industries 336413 and 331419? The evidence presented in Section III.B that *LT1* is an important driver of *PC1* suggests that we look at what explains correlated liquidity flow and why such trades have a price impact. First, measurement error may cause some marketwide private information trades to be counted as liquidity trades. Second, as argued before, we do not believe that indexing plays a major role in explaining the presence of comovement and, hence, of liquidity trades. Third, DeLong et al. (1990) suggest that correlated noise trading risk can generate long-run price effects. DeLong et al. (1990) argue that noise traders' unpredictable beliefs create a risk in the price of the asset that deters rational arbitrageurs from eliminating price deviations from fundamental values. In their model noise traders earn higher expected returns, a situation that predicts the ability of *PC1* to forecast returns. Campbell and Kyle (1993) also have

a model in which noise trading is correlated with fundamentals. The separate roles of *MPI* and *PC1* and their joint properties is an interesting topic for future research.

As a final robustness check we wish to discount the possibility that in the currency regressions joint significance is the result of some serial correlation in currency returns (albeit undetected) that are captured by the lags of *MPI*. For this, we perform the same regressions for each currency on lags of *MPI* and on lag values of the dependent variable. This amounts to the estimation of several ARMAX ($p, 0, 10$) specifications where lag selection for the AR terms is conducted using Akaike's Information Criterion or set to one whenever this delivers zero lags. The analysis is only done for contemporaneous and 1-month-ahead currency returns because the overlapping observations-induced autocorrelation causes an orthogonality violation for the regressions on 2-month-ahead currency returns.¹⁶ The results (available upon request) show that adding autoregressive lags does not change the Wald tests on the explanatory power of the contemporaneous and lagged *MPI*. Indeed, only for industry 336412 and for the simple and excess returns on the British pound do we find that at the optimal AR lag length the current and lagged *MPI* ceases to be significant.

F. An Alternative Approach to Estimating *MPI*?

There is a significant computational burden involved in estimating *MPI*. This derives from the many parameters that need to be estimated, but also from the highly nonlinear objective function. The latter requires thoroughness in searching for a global maximum via a complete search of the parameter space. Moreover, the estimation becomes slower as we add more firms.

In this subsection, we study a potentially simpler way of estimating *MPI* that sidesteps the need to do a joint estimation of the likelihood function. We know from Table II and Section III.E above that the first principal component in industry order flow *PC1* is not a good substitute for *MPI*. Can we do a better job using the first principal component in order flow due to firm-specific trading estimated instead of using *PC1* (and assuming $\mu_i^a = 0$)? To implement this alternative, consider estimating the model in EKOP. This amounts to maximizing for each firm in an industry the likelihood function

$$\begin{aligned}
 l(\{S_i, B_i\}) = \prod_{n=1}^N & \left[\alpha_i (1 - \delta_i) e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-(\mu_i^f + \varepsilon_i^b)} \frac{(\mu_i^f + \varepsilon_i^b)^{B_{in}}}{B_{in}!} \right. \\
 & + \alpha_i \delta_i e^{-(\mu_i^f + \varepsilon_i^s)} \frac{(\mu_i^f + \varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!} \\
 & \left. + (1 - \alpha_i) e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!} \right]. \quad (17)
 \end{aligned}$$

¹⁶ The lack of obvious instruments for currency returns excludes the possibility of using 2SLS: In most instances currency returns are i.i.d., so lagged returns are necessarily poor instruments.

This problem is easy and fast to estimate. Once estimates for the parameters are obtained, we compute the following measure of trades due to firm-specific private information:

$$FPI_i = \hat{\alpha}_i(1 - 2\hat{\delta}_i)\hat{\mu}_i^f. \quad (18)$$

The procedure yields a time series of FPI_i for each firm i . We then estimate the first principal component out of the various FPI_i in an industry, which we label MPI' . The variable MPI' captures co-movement in the measures of firm-specific private information and is thus a natural candidate for an alternative measure of marketwide private information. One problem with computing MPI' is that throughout the sample firms enter and exit, and some firms may be temporarily excluded from our sample due to low liquidity. We deal with this problem by estimating the first principal component in the FPI s over each different block of data.

To evaluate MPI' , we first compute the percentage of explained variation in MPI —our original measure—explained by the new measure MPI' , that is, the R^2 in a regression of MPI on MPI' . Except for industry 336413, the regression F -tests are significant (p -values below 5%). However, the R^2 s are quite low; the most that MPI' explains of MPI is 18.5% for industry 333132. For industries 336411, 336412, and 331419, the R^2 s are 14%, 15.5%, and 4%, respectively (un-tabulated). Second, we use MPI' and repeat the exercises leading to Tables III through VIII. The results (available upon request) are broadly consistent with MPI' being at best a noisier measure of marketwide private information than is MPI . Most equity-return forecasting regressions fail to deliver estimated parameters whose sum is statistically positive. Also, in industry 336412—where the equity-return regressions show the best results— MPI' has no forecasting power over the respective currency-basket return. In contrast, MPI' has some success forecasting the currency returns in industries 336413 and 331419, but these results are at odds with the ability of MPI' to forecast equity returns in the same industries.

IV. Conclusion

This paper presents a model of equity trading with informed and uninformed investors, where informed investors act upon firm-specific and marketwide private information. Using the model we estimate the component of order flow that is due to marketwide private information.

Marketwide private information displays positive but low or statistically insignificant correlation with the first principal component in order flow, which suggests that marketwide private information is not well captured by a simple statistical factor of order flow. Instead, using our structural model, we show that estimated liquidity trades display a significant positive association with the first principal component in order flow.

We conduct our analysis to test whether marketwide private information permeates both the equity and currency markets. We find that marketwide private information obtained from stock market order flow data forecasts industry

stock returns and foreign exchange returns, consistent with the label of marketwide information. Our findings are consistent with the model of exchange rate determination in Evans and Lyons (2004a) and are important to many models of asset pricing that assume the existence of private information at an index level.

Asking whether any information exists that can help market makers (and econometricians) better understand the nature of the information that is about to hit the equity market is an interesting question. For example, macro news announcements and firm earnings announcements might be used to model the parameters that dictate the arrival of information. An alternative would be to use currency order flow: If *MPI* contains relevant information about currency markets and if trading on that information first occurs in such markets because they are more liquid, then conditioning on currency order flow improves the model's fit.

Appendix A: Sample Selection

This appendix discusses in detail the selection of industries and firms included in the analysis. We start with the 90 manufacturing six-digit NAICS industries that have the highest levels of exports in each of the years from 1997 to 2003. Data for exports are the FAS value of domestic exports obtained from the U.S. International Trade Commission database, available at <http://dataweb.usitc.gov/>. Our focus on manufacturing industries (i.e., NAICS 31-33) is without loss of generality because manufacturing represents roughly 90% of all U.S. exports. These data are merged with shipment data from the U.S. Census Bureau Annual Survey of Manufactures, available at <http://www.census.gov/mcd/>. Before 1997, industries were organized according to SIC codes and due to an imperfect bridge between NAICS and SIC the U.S. Census Bureau has identified a significant understatement of shipments prior to 1997 using the NAICS system. Therefore, we base our industry selection only on data from 1997 to 2003.

We compute the ratio of domestic exports to total shipments for each of the years and rank the industries in each year according to this ratio. Table AI presents the top 30 industries of each year from highest to lowest export-to-shipment ratio. The industries that end up being selected after we apply our filters are designated with an asterisk.

To guarantee continued top foreign exposure, we drop all the industries that do not rank among the top 30 in any one of the years from 1997 to 2003. This leaves us with 20 industries identified in column 9 of Table AI ("Continuous Top 30 Exporters"), listed according to their ranking in 2003. Next, because our quotes data go back to 1993 when SIC codes were used, agreement between SIC and NAICS is required to guarantee that firms are treated consistently before and after 1997 and that export data can be used to identify major currency exposures and to construct currency baskets. For this reason we drop the "almost comparable" and the "not comparable" industries (see <http://www.census.gov/epcd/naics02/N2SIC31B.HTM#N315>), resulting in

Table AI
Selection of Industries with Foreign Exposure and Comparability between SIC and NAICS Codes

The table shows the top 30 manufacturing exporters of every year from 1997 to 2003 based on the ratio of exports to shipments. It also shows the NAICS industries that continuously ranked in the top 30 and had a complete bridge with the SIC codes. The industries used in the analysis are designated with an asterisk (*).

Position Ranked	Years							Continuous Top 30 Exporters	NAICS and SIC Comparable	
	1997	1998	1999	2000	2001	2002	2003		2003 Rank	Average Rank
1	331419*	331419*	331419*	331419*	331419*	315221	315221	315221		331419*
2	334119	334119	334119	334119	334119	331419*	331419*	331419*	331419*	
3	333999	333132*	333999	333314	333314	334119	334119	334119		
4	333132*	333999	325311	333999	333999	333132*	313249			
5	333314	325311	333132*	333132*	315221	333314	333314	333314		
6	315999	315999	333314	323119	333132*	333999	333999	333999		333132*
7	323119	333314	323119	315221	323119	323119	333132*	333132*	333132*	
8	325311	323119	315999	325311	334513	315999	323119	323119		
9	333611	315221	336413*	315999	336413*	313249	315999	315999		325311
10	336413*	336411*	334513	334513	325311	313221	334513	334513	334513	336413*
11	335999	333611	333611	336413*	313221	336413*	336413*	336413*	336413*	
12	336399	336413*	336412*	336412*	315999	334513	313221	313221		336412*
13	334513	334310	315221	335999	313249	325311	336412*	336412*	336412*	
14	322110	334513	336411*	334515	334419	336412*	325311	325311	325311	
15	334419	336399	335999	313221	336412*	322110	334413	334413	334413	322110
16	336411*	335999	322110	322110	334413	334413	315223			
17	334310	322110	336399	334413	335999	334515	322110	322110	322110	
18	333295	336412*	334413	334419	334515	336399	336120			334413
19	336120	334419	313221	332999	322110	316110	325411			

(continued)

Table AI—Continued

Position Ranked	Years							Continuous Top 30 Exporters	NAICS and SIC Comparable	
	1997	1998	1999	2000	2001	2002	2003		2003 Rank	Average Rank
20	315221	313221	334515	336399	334310	335999	334515	334515		336411*
21	336412*	334515	334419	334310	336399	336411*	316110	316110		
22	313221	334413	334310	333611	336120	334310	325192	325192		
23	334515	316110	311223	333295	336411*	336120	336399	336399		
24	334413	311223	333295	336411*	333611	334419	335999	335999		
25	311212	311212	311212	313249	316110	323117	311212	311212		
26	323117	336120	332999	336120	325411	333295	323117	323117		
27	333120	323117	334510	311212	333295	311223	336411*	336411*	336411*	
28	311223	325199	316110	334510	325188	325188	313320	313320		
29	325414	333295	333111	333120	334510	315223	325188	325188		
30	333111	332999	323117	325188	332999	333618	334419	334419		

Table AII
Selection of Industries and Data Availability

The table shows the implications for the number of firms in each industry as a result of applying our selection filters. The industries used in the analysis are designated with an asterisk (*).

NAICS Average Rank 1997–2003	Number of NYSE Listed Firms	Foreign Incorporated or Missing Data	Firms That Fail to Meet the Liquidity Criterion	Number of Available Firms
331419*	2	2	0	4
333132*	16	5	0	11
325311	7	5	1	1
336413*	11	3	0	8
334513	7	2	0	5
336412*	6	0	0	6
322110	4	1	0	3
334413	27	13	1	13
336411*	7	2	1	4

a loss of 11 industries. Column 10 of Table AI (“NAICS & SIC Comparable, 2003 Rank”) lists the nine remaining industries according to their rank in 2003, and column 11 (“NAICS & SIC Comparable, Average Rank”) lists the same industries according to their average rank throughout the sample.

Table AII lists the nine industries from Table AI, which include the five industries we study (designated with an asterisk). Column 2 of Table AII indicates the number of firms that in total belonged to each of the industries at some point during our sample and were listed in the NYSE as indicated by Compustat. Compustat only has information about the latest exchange of listing of a company’s stock, so we cross-check their information with the SEC filings on EDGAR (<http://www.sec.gov/edgar>). Column 3 lists the number of foreign firms or firms with insufficient data in each industry (i.e., mostly the latter refer to inactive firms during our sample). Column 4 lists the number of firms that do not meet the liquidity criterion over the full sample, and column 5 gives the maximum number of firms available at any month for estimation by industry (column 5 = column 2 – column 3 – column 4). Industries NAICS 325311 and 322110 are excluded from the analysis for having too few firms to include in the estimation of marketwide private information. Note that four firms is the smallest number of firms in any industry we study in the paper. Industry NAICS 334413 is excluded because it has too many firms. As discussed in the main text, the estimation cannot evaluate the likelihood function when the number of buy and sell orders is large (Vega (2006) also encounters the same problem) and when the number of firms is also large. Industry NAICS 334513 is dropped because it ranked last across all the remaining six industries in terms of volume of exports (with 20% less exports than the next-to-last industry NAICS 333132), leaving us with five industries. Finally, INCO and WMC/Alumina in NAICS 331419 and Doncasters PLC in NAICS 336412 have foreign incorporation, but are included in the analysis due to the size of their

operations in the U.S. and/or the fact that the U.S. dollar is the currency of denomination in the industry.

Appendix B: Currency Exposures

This appendix provides further information on the currency exposure of the firms in our study as mentioned in their annual reports and 10-K forms. To conserve space we only report on information regarding representative years.

B1. Oil and Gas Field Machinery and Equipment Manufacturing: NAICS 333132

In 2003 Baker Hughes reports entering into foreign currency forwards to partially hedge exposure to currency fluctuations in currencies such as the British pound, the Norwegian krone, the euro, the Brazilian real, and the Argentine peso. Baker Hughes also acknowledges exposure in previous years to the Canadian dollar and the Indonesian rupiah. Weatherford International's functional currency for international operations is the applicable local currency. However, it has a natural hedge from local expenses of foreign operations. Its 2001 annual report indicates that approximately 27% of net assets are impacted by changes in foreign currencies relative to the U.S. dollar. Cooper Cameron has production facilities located in the United Kingdom and other European and Asian countries. The firm's profitability is eroded when the U.S. dollar weakens against the British pound, the euro, and other currencies. Indeed, Cooper Cameron was negatively impacted during 2003 as a result of the weakening U.S. dollar and may be further negatively impacted if the U.S. dollar continues to weaken. Varco's 2002 annual report says that the losses occurred in the second quarter of 2002 were due mostly to the weakening of the U.S. dollar against the euro and the British pound. Similarly, FMC Technologies reports exposure to the euro, the British pound, the Norwegian krone, and the Japanese yen, among others. National-Oilwell uses the local currency for its operations in Canada, the U.K., Germany, and Australia and reports that it did not engage or plan to engage in any significant hedging. NATCO Group and Grant Prideco also report using the Canadian dollar as the functional currency for their operations in Canada. NATCO is exposed to fluctuations in the cross-rate British pound/euro, whereas Grant Prideco is exposed to Venezuelan and Chinese currencies.

B2. Aircraft Manufacturing: NAICS 336411

In their annual reports, the firms in this industry acknowledge exposure to the Japanese yen, Australian dollar, Canadian dollar, and several European currencies. For instance, Boeing's 2003 annual report acknowledges foreign currency exposure as a result of suppliers and subcontractors located in Europe, even though most operations are in the United States, Canada, and Australia. Although Boeing's foreign operations only accounted for 2% of total sales, 40% of its revenue came from foreign clients. As discussed in the main text, even when

these clients are invoiced in U.S. dollar, Boeing's operating exposure remains; its competitiveness is affected by U.S. dollar movements. As with Boeing, Grumman Corp. also hedges most of its foreign currency transactions exposure. However, it does not hedge translation exposure resulting from operations abroad.

*B3. Aircraft Engine and Engine Parts
Manufacturing: NAICS 336412*

Sequa reports primary foreign currency exposure to the British pound/U.S. dollar and the British pound/euro from its U.K. firm Warwick International. UNC Inc. has significant foreign operations but uses natural hedges and financial hedges to eliminate most of its transactions exposure. However, it does not hedge any of its translation exposure from foreign currency net assets. United Technologies had at one point over 50% of its revenue coming from foreign markets, namely, Europe and Asia-Pacific. Doncasters PLC uses the local currency as its functional currency in its foreign operations, reporting exposures to British pound/U.S. dollar and the British pound/euro exchange rates. During our sample period, Howmet International had operations in France, the United Kingdom, Canada, and Japan. It also had forward exchange rate contracts partially hedging exposure to the British pound, the French franc, and the Japanese yen.

*B4. Other Aircraft Parts and Auxiliary Equipment
Manufacturing: NAICS 336413*

Most of the currencies that represent the export markets of this industry were explicitly referred to by firms in their annual reports and by other financial reports. For example, Honeywell International, Goodrich Corp., and others report principal exposures to the British pound, the euro (the German mark before it), and the Canadian dollar. Honeywell hedges most of its transactions exposure, but does not hedge exposure resulting from translation of foreign currency cash flows and net assets. Sundstrand Corp. acknowledges hedging most of its exposure to fluctuations in foreign currencies for transactions denominated primarily in the British pound, French franc, and Singapore dollar. Sundstrand does not hedge any of its translation exposure. Likewise, Rockwell Collins Inc. faces significant exposure to fluctuations in exchange rates though it actively manages most of it, except that associated with translation exposure.

*B5. Primary Smelting and Refining of Nonferrous Metal (except Copper
and Aluminum): NAICS 331419*

Reading through the company's annual reports, we observe reported exposures to the currencies representing the major export markets and other currencies. For example, in 2001 Tremont's annual report indicates that earnings were primarily affected by fluctuations in the value of the U.S. dollar relative to the euro, the Canadian dollar, the Norwegian kroner, and the British pound. In

2001 approximately 40% of TIMET's, Tremont's main subsidiary, sales revenue originated in Europe, 60% of which was denominated in currencies other than the U.S. dollar, mainly the British pound and European currencies now in the euro area. WHX reports currency exposures related to anticipated revenues and operating costs and commitments for capital expenditures in foreign currencies, but says it does not hedge these exposures. INCO says in its 2003 annual report that notwithstanding the use of foreign currency forwards on the Canadian dollar, the euro, the Australian dollar, and the Indonesian rupiah, changes in exchange rates can have a material impact on future earnings and cash flows. This exposure arises primarily from costs of foreign operations denominated in local currencies. WMC reported consistently not hedging currency risk from fluctuations in the Australian dollar/U.S. dollar rate within Alumina Limited or AWAC.

Appendix C: Estimated Correlation between LT and MPI

This appendix presents a discussion of the estimated negative correlation between liquidity trades and trades driven by private information. We show that estimates of μ_i^a (or μ_i^f) are negatively correlated with estimates of ε_i^b and ε_i^s , inducing a correlation between informed trading and liquidity trading. Intuitively, as the model tries to match the total number of buy or sell orders, the more orders explained by marketwide-driven private information trading, the fewer attributed to liquidity trading.

To formalize this intuition we make the following assumptions. First, we assume that the underlying true distributional parameters in each industry are constant over time. The sample covariance between the estimates $\widehat{LT}_i$ and $\widehat{MPI}_i$ for firm i is $\frac{1}{T-1} \sum_t (\widehat{LT}_{it} - \overline{\widehat{LT}}_i) \times (\widehat{MPI}_{it} - \overline{\widehat{MPI}}_i)$, where $\overline{\widehat{LT}}_i = \frac{1}{T} \sum_t \widehat{LT}_{it}$ and $\overline{\widehat{MPI}}_i = \frac{1}{T} \sum_t \widehat{MPI}_{it}$. With random samples of buy and sell orders drawn over time, and with the assumption of constant distributional parameters, $\overline{\widehat{LT}}_i \rightarrow LT_i$ and $\overline{\widehat{MPI}}_i \rightarrow MPI_i$ under standard regularity conditions, where LT_i and MPI_i are evaluated at the true population parameters. Therefore, for large T , we are interested in how small sample estimation errors in $\widehat{LT}_{it}$ and $\widehat{MPI}_{it}$ comove. Notice that we have shifted the analysis from the covariance between industry liquidity trading and marketwide private information to firm i 's liquidity trading and marketwide private information trades. This is inconsequential if industries are composed of similar firms, which we assume.

Second, we assume that there is no firm-specific private information, that is, $\alpha_i = 0$ for all i : All trades are driven by either liquidity trades or marketwide private information trades. This assumption is made for simplicity only. The argument below applies if we assume instead that only firm-specific private information exists.¹⁷ Finally, we assume that the estimation applies to the parameters μ_i^a , ε_i^b , and ε_i^s for all i , and that θ and ρ are known.

¹⁷ We have not been able to obtain conditions that apply when both μ_i^a and μ_i^f coexist. By continuity, our argument applies as either μ_i^a or μ_i^f approach zero. However, we have constructed examples where the correlation between estimated MPI and estimated LT is positive.

Under these assumptions, equations (1) to (3) in the main text become

$$\begin{aligned} & \theta(1 - \rho)\Pi_{i=1}^I l_G(\{S_i, B_i\}) \\ &= \theta(1 - \rho)\Pi_{i=1}^I \left[e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!} e^{-\mu_i^a} \left(\frac{\mu_i^a}{\varepsilon_i^b} + 1 \right)^{B_{in}} \right], \end{aligned} \quad (C1)$$

$$\theta\rho\Pi_{i=1}^I l_B(\{S_i, B_i\}) = \theta\rho\Pi_{i=1}^I \left[e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!} e^{-\mu_i^a} \left(\frac{\mu_i^a}{\varepsilon_i^s} + 1 \right)^{S_{in}} \right], \quad (C2)$$

and

$$(1 - \theta)\Pi_{i=1}^I l_0(\{S_i, B_i\}) = (1 - \theta)\Pi_{i=1}^I \left[e^{-\varepsilon_i^s} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} e^{-\varepsilon_i^b} \frac{(\varepsilon_i^b)^{B_{in}}}{B_{in}!} \right]. \quad (C3)$$

The log-likelihood of observing $I \times N$ buy and sell orders $\{B_{in}, S_{in}\}$, $\log L((S_{in}, B_{in})_{in})$, is

$$\sum_{n=1}^N \sum_{i=1}^I \left[-(\varepsilon_i^s + \varepsilon_i^b) + S_{in} \log \varepsilon_i^s + B_{in} \log \varepsilon_i^b - \log(S_{in}! B_{in}!) \right] + \sum_{n=1}^N \log D_{in}, \quad (C4)$$

where

$$D_{in} \equiv \theta(1 - \rho)\Pi_{i=1}^I e^{-\mu_i^a} \left(\frac{\mu_i^a}{\varepsilon_i^b} + 1 \right)^{B_{in}} + \theta\rho\Pi_{i=1}^I e^{-\mu_i^a} \left(\frac{\mu_i^a}{\varepsilon_i^s} + 1 \right)^{S_{in}} + 1 - \theta. \quad (C5)$$

Denote the average number of buy orders for firm i by $\bar{B}_i = \frac{1}{N} \sum_{n=1}^N B_{in}$ and the average number of sell orders by $\bar{S}_i = \frac{1}{N} \sum_{n=1}^N S_{in}$. The first-order conditions for ε_i^b and ε_i^s , evaluated at the MLE, can be expressed as

$$\hat{\varepsilon}_i^b = \bar{B}_i - \frac{1}{N} \sum_{n=1}^N B_{in} \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \hat{\varepsilon}_i^b} D_{in}^{-1} \theta(1 - \rho)\Pi_{j=1}^I e^{-\hat{\mu}_j^a} \left(\frac{\hat{\mu}_j^a}{\hat{\varepsilon}_j^b} + 1 \right)^{B_{jt}}, \quad (C6)$$

$$\hat{\varepsilon}_i^s = \bar{S}_i - \frac{1}{N} \sum_{n=1}^N S_{in} \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \hat{\varepsilon}_i^s} D_{in}^{-1} \theta\rho\Pi_{j=1}^I e^{-\hat{\mu}_j^a} \left(\frac{\hat{\mu}_j^a}{\hat{\varepsilon}_j^s} + 1 \right)^{S_{jt}}. \quad (C7)$$

We omit the first-order conditions for the μ_i^a because these are uninteresting for our argument.

Under the additional restriction of no marketwide private information, that is $\mu_i^a = 0$,

$$\hat{\varepsilon}_i^b = \bar{B}_i, \quad \hat{\varepsilon}_i^s = \bar{S}_i, \quad (C8)$$

and the model captures total order flow perfectly by correctly assigning its entirety to liquidity trades.

To derive an approximate solution absent the restrictions on μ_i^a assume that, for all j and all n ,

$$\left(\frac{\hat{\mu}_j^a}{\hat{\varepsilon}_j^b} + 1\right)^{B_{jn}} \approx e^{\hat{\mu}_j^a}, \quad \left(\frac{\hat{\mu}_j^a}{\hat{\varepsilon}_j^s} + 1\right)^{S_{jn}} \approx e^{\hat{\mu}_j^a}. \quad (\text{C9})$$

These would be exact if $\hat{\varepsilon}_j^b = B_{jn} \rightarrow \infty$ and $\hat{\varepsilon}_j^s = S_{jn} \rightarrow \infty$. They are a reasonable approximation when liquidity trading is large relative to all other sources of trading and given that large buy and sell orders push up the estimated parameters ε_j^b and ε_j^s (see (C8)). With this approximation the first-order conditions simplify to

$$\hat{\varepsilon}_i^b \simeq \bar{B}_i - \bar{B}_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \hat{\varepsilon}_i^b} \theta (1 - \rho), \quad (\text{C10})$$

$$\hat{\varepsilon}_i^s \simeq \bar{S}_i - \bar{S}_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \hat{\varepsilon}_i^s} \theta \rho. \quad (\text{C11})$$

We next show that estimated $LT_i = \varepsilon_i^b - \varepsilon_i^s$ varies negatively with informed trading. When $\alpha_i = 0$ informed trading becomes $MPI_i = (1 - 2\rho)\mu_i^a$. Consider two extreme cases for parameter ρ .

Case 1: *Marketwide private information is always good news*, $\rho = 0$. Then $\hat{\varepsilon}_i^s = \bar{S}_i$, $MPI_i = \mu_i^a$, and

$$\begin{aligned} \frac{\partial \widehat{LT}_i}{\partial \widehat{MPI}_i} &= \frac{\partial \hat{\varepsilon}_i^b}{\partial \hat{\mu}_i^a} \\ &= -\frac{\hat{\varepsilon}_i^b - \bar{B}_i(1 - \theta)}{\hat{\mu}_i^a - \bar{B}_i + 2\hat{\varepsilon}_i^b} < 0, \end{aligned} \quad (\text{C12})$$

where the partial derivative is obtained from (C10) and takes θ as fixed. The inequality arises because $\hat{\varepsilon}_i^b[(\hat{\mu}_i^a - \bar{B}_i) + \hat{\varepsilon}_i^b] = \bar{B}_i\hat{\mu}_i^a(1 - \theta) > 0$ after rewriting (C10) and, again using (C10),

$$\hat{\varepsilon}_i^b \simeq \bar{B}_i - \bar{B}_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \hat{\varepsilon}_i^b} \theta > \bar{B}_i - \bar{B}_i \theta > 0. \quad (\text{C13})$$

Intuitively, the level of liquidity sell orders only depends on the average number of sell orders. Thus, estimation error in $\hat{\mu}_i^a$ can only affect estimates of average liquidity buy orders $\hat{\varepsilon}_i^b$ and does so in a negative way. As the model tries to estimate parameters in order to match total buy orders, a positive small sample estimation error in $\hat{\mu}_i^a$ is associated with a negative small sample estimation error in estimated LT_i .

Case 2: *All marketwide private information is bad news*, $\rho = 1$. Then $\hat{\varepsilon}_i^b = \bar{B}_i$, $MPI_i = -\mu_i^a$, and

$$\begin{aligned}
\frac{\partial \widehat{LT}_i}{\partial \widehat{MPI}_i} &= \frac{\partial \hat{\varepsilon}_i^s}{\partial \hat{\mu}_i^a} \\
&= -\frac{\hat{\varepsilon}_i^s - \bar{S}_i(1-\theta)}{\hat{\mu}_i^a - \bar{S}_i + 2\hat{\varepsilon}_i^s} < 0,
\end{aligned} \tag{C14}$$

where the partial derivative is obtained from (C11) and again takes θ as constant. The sign of the partial derivative is justified because $\hat{\varepsilon}_i^s[(\mu_i^a - \bar{S}_i) + \hat{\varepsilon}_i^s] = \bar{S}_i\hat{\mu}_i^a(1-\theta) > 0$ after rewriting (C11), and, also using (C11),

$$\hat{\varepsilon}_i^s \simeq \bar{S}_i - \bar{S}_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \hat{\varepsilon}_i^s} \theta > \bar{S}_i - \bar{S}_i \theta > 0. \tag{C15}$$

The intuition for the negative association between $\widehat{LT}_i$ and $\widehat{MPI}_i$ relies on the dependence of liquidity buy orders on the average number of buy orders. An increase in $\hat{\mu}_i^a$ only affects estimates of average liquidity sell orders $\hat{\varepsilon}_i^s$ and does so in a negative way. However, estimates of $\widehat{MPI}_i = -\mu_i^a$ also move in the same direction as $\hat{\varepsilon}_i^s$. As the model estimates parameters to match total sell orders, any positive estimation error in $\hat{\mu}_i^a$ that lowers the estimate of $\widehat{MPI}_i$ increases estimated $\widehat{LT}_i$.

General case: $\rho \in [0, 1]$. Using (C10)–(C11) we obtain

$$\frac{\partial \widehat{LT}_i}{\partial \hat{\mu}_i^a} = -\frac{\hat{\varepsilon}_i^b - \bar{B}_i(1-\theta(1-\rho))}{\hat{\mu}_i^a - \bar{B}_i + 2\hat{\varepsilon}_i^b} + \frac{\hat{\varepsilon}_i^s - \bar{S}_i(1-\theta\rho)}{\hat{\mu}_i^a - \bar{S}_i + 2\hat{\varepsilon}_i^s}. \tag{C16}$$

While it is difficult to sign this derivative in general, for low values of $\hat{\mu}_i^a$ we can go a long way. If $\hat{\mu}_i^a = 0$, then $\hat{\varepsilon}_i^s = \bar{S}_i$ and $\hat{\varepsilon}_i^b = \bar{B}_i$ (see (C8)), implying that

$$\left. \frac{\partial \widehat{LT}_i}{\partial \hat{\mu}_i^a} \right|_{\hat{\mu}_i^a=0} = -\theta(1-2\rho). \tag{C17}$$

Given that $\frac{\partial \widehat{MPI}_i}{\partial \hat{\mu}_i^a} = 1 - 2\rho$ we have

$$\left. \frac{\partial \widehat{LT}_i}{\partial \widehat{MPI}_i} \right|_{\hat{\mu}_i^a=0} = -\theta < 0. \tag{C18}$$

Therefore, we obtain a negative correlation.

REFERENCES

- Albuquerque, Rui, Greg Bauer, and Martin Schneider, 2005, Global private information in international equity markets, Working paper, Boston University.
- Bacchetta, Philippe, and Eric van Wincoop, 2006, Can information heterogeneity explain the exchange rate determination puzzle? *American Economic Review* 96, 552–576.

- Barclay, Michael J., Robert H. Litzenberger, and Jerold B. Warner, 1990, Private information, trading volume, and stock-return variances, *Review of Financial Studies* 3, 233–253.
- Bauer, Gregory H., and Clara Vega, 2007, The monetary origins of asymmetric information in international equity markets, *Journal of International Money and Finance* (forthcoming).
- Brennan, Michael J., and H. Henry Cao, 1997, International portfolio investment flows, *Journal of Finance* 52, 1851–1880.
- Caballe, Jordi, and Murugappa Krishnan, 1994, Imperfect competition in a multi-security market with risk neutrality, *Econometrica* 62, 695–704.
- Campbell, John, and Albert S. Kyle, 1993, Smart money, noise trading and stock price behavior, *Review of Economic Studies* 60, 1–34.
- Chan, Kalok, 1993, Imperfect information and cross-autocorrelation among stock prices, *Journal of Finance* 48, 1211–1230.
- Chan, Kalok, and Allaudeen Hameed, 2006, Stock price synchronicity and analyst coverage in emerging markets, *Journal of Financial Economics* 80, 115–147.
- DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.
- Dufour, J.-M., and Olivier Torr  s, 1998, Union-intersection and sample-split methods in econometrics with applications to MA and SURE models, in Griles, D. and Aman Ullah, eds.: *Handbook of Applied Economic Statistics* (Marcel Dekker, New York).
- Easley, David, Robert F. Engle, Maureen O'Hara, and Liuren Wu, 2002, Time-varying arrival rates of informed and uninformed trades, Working paper, Cornell University.
- Easley, David, and Maureen O'Hara, 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577–605.
- Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997, One day in the life of a very common stock, *Review of Financial Studies* 10, 805–835.
- Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman, 1996, Liquidity, information, and infrequently traded stocks, *Journal of Finance* 51, 1405–1436.
- Ellis, Katrina, Roni Michaely, and Maureen O'Hara, 2000, The accuracy of trade classification rules: Evidence from Nasdaq, *Journal of Financial and Quantitative Analysis* 35, 529–551.
- Evans, Martin D. D., and Richard K. Lyons, 2002, Order flow and exchange rate dynamics, *Journal of Political Economy* 110, 170–180.
- Evans, Martin D. D., and Richard K. Lyons, 2004a, A new micro model of exchange rate dynamics, Working paper, UC Berkeley.
- Evans, Martin D. D., and Richard K. Lyons, 2004b, Exchange rate fundamentals and order flow, Working paper, UC Berkeley.
- Francis, Bill B., Iftekhar Hasan, and Delroy M. Hunter, 2006, Dynamic relations between international equity and currency markets: The role of currency order flow, *Journal of Business* 79, 219–258.
- Gehrig, Thomas, 1993, An information based explanation of the domestic bias in international equity investment, *Scandinavian Journal of Economics* 95, 97–109.
- Gertler, Mark, and Simon Gilchrist, 1994, Monetary policy, business cycles, and the behavior of small manufacturing firms, *Quarterly Journal of Economics* 109, 309–340.
- Goldberg, Linda S., and C  dric Tille, 2004, Vehicle currency use in international trade, Working paper, Federal Reserve Bank of New York.
- Hansen, Lars, and Robert J. Hodrick, 1980, Forward exchange rates as optimal predictors of future spot exchange rates: An econometric analysis, *Journal of Political Economy* 88, 829–853.
- Harford, Jarrad, and Aditya Kaul, 2005, Correlated order flow: Pervasiveness, sources and pricing effects, *Journal of Financial and Quantitative Analysis* 40, 29–55.
- Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179–207.
- Hasbrouck, Joel, and Duane J. Seppi, 2001, Common factors in prices, order flows, and liquidity, *Journal of Financial Economics* 59, 383–411.
- Kumar, Praveen, and Duane J. Seppi, 1994, Information and index arbitrage, *Journal of Business* 67, 481–509.

- Lee, Charles M. C., and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from TORQ data, *Journal of Financial Markets* 3, 83–111.
- Lee, Charles M. C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.
- Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.
- Lyons, Richard, 1997, A simultaneous trade model of the foreign exchange hot potato, *Journal of International Economics* 42, 275–298.
- Newey, Whitney K., and Daniel McFadden, 1994, Large sample estimation and hypothesis testing, in R. Engle and D. McFadden, eds.: *Handbook of Econometrics*, Volume 4 (North-Holland, Amsterdam).
- Odders-White, Elizabeth R., 2000, On the occurrence and consequences of inaccurate trade classification, *Journal of Financial Markets* 3, 259–286.
- Pagan, Adrian, 1984, Econometric issues in the analysis of regressions with generated regressors, *International Economic Review* 25, 221–246.
- Rime, Dagfinn, 2001, Private or public information in foreign exchange markets? An empirical analysis, Working paper, Central Bank of Norway.
- Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.
- Subrahmanyam, Avanidhar, 1991, A theory of trading in stock index futures, *Review of Financial Studies* 4, 17–51.
- Vega, Clara, 2006, Stock price reaction to public and private information, *Journal of Financial Economics* 82, 103–133.
- Vives, Xavier, 1993, How fast do rational agents learn? *Review of Economic Studies* 60, 329–347.
- Yu, Lei, 2005, Basket securities, price formation, and informational efficiency, Working paper, University of Notre Dame.