Marketwide Private Information in Stocks:
Forecasting Currency Returns*

Rui Albuquerque†  Eva de Francisco‡  Luis Marques§

February 21, 2007

Abstract

We present a model of equity trading with informed and uninformed investors where informed investors act upon firm-specific private information and marketwide private information. The model is used to structurally identify the component of order flow that is due to marketwide private information. Estimated trades driven by marketwide private information display very little or no correlation with the first principal component in order flow. This finding implies that a simple statistical factor of order flow is a poor measure of marketwide private information. Moreover, the model suggests that the previously documented co-movement in order flow captures mostly variation in liquidity trades. Marketwide private information obtained from equity market data forecasts industry stock returns. It also forecasts foreign exchange returns consistent with Evans and Lyons’ (2004a) model of exchange rate determination.

JEL Classification: F31, G11, G14.
Keywords: Marketwide private information, firm-specific private information, order flow, principal components, currency returns, equity returns.

*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 Instituto de Empresa Business School in Madrid, the University of Rochester and the Federal Reserve Board for comments. The paper has greatly benefitted from comments by an anonymous referee and an associate editor. The usual disclaimer applies.

†Albuquerque is at the Boston University School of Management and CEPR. Address at BU: 595 Commonwealth Avenue, Boston, MA 02215. Phone: 617-277-2342. Email: ralbuque@bu.edu.

‡de Francisco is affiliated with the Department of Economics at Towson University, 8000 York Road, Towson, MD 21252. Email: edefrancisco@towson.edu.

§Marques is affiliated with the Department of Economics at the University of Rochester, Harkness Hall, Rochester, NY 14627. Email: lmrq@troi.cc.rochester.edu.
1 Introduction

Markets aggregate dispersed information from economic agents and impound it into prices. This information originates from public sources in the form of company reports and official statistics, or from private investor sources in their use of proprietary models, expertise and insider knowledge. Its content is either asset-specific or aggregate in nature. 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 as they fluctuate with industry or economy wide business conditions, or about discount rates as they move with the economy’s riskless interest rate and aggregate risk premium. In contrast, *firm-specific private information* is idiosyncratic in nature and useless in 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 with the aim of structurally identifying marketwide private information from firm-specific private information and liquidity trades. The model generalizes Easley et al. (EKOP, 1996) by allowing for trading in multiple stocks and in private information at two levels, firm-specific and marketwide. Our identifying assumption is that marketwide private information generates trading in several stocks simultaneously. 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 offset each other requires an ex-ante choice of weights to each information signal: we assume that informed investors discard their marketwide private signals, 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 preformed to validate the estimation results.

¹These 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.
The paper’s second contribution is to study the properties of marketwide private information. Estimated marketwide private information has three main properties. First, in each industry 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 later fact implies that marketwide private information cannot be replicated by a simple statistical procedure like a principal component analysis. Instead, our estimate of liquidity trades correlates strongly with the principal component in order flow in the data, 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 the criticism in Harford and Kaul (2005) that co-movement in order flow is driven by firms sharing the same index.

The second property of marketwide private information is that it forecasts the stock returns of the firms in the industries we study. Marketwide private information is able to forecast returns up to two 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 non-informative inventory or liquidity effects.

The third property of marketwide private information is that it forecasts currency returns. This finding constitutes evidence in favor of Evans and Lyons (2004a) and gives additional credence to our label of marketwide private information. We regress changes in industry-specific currency baskets on marketwide private information. We repeat the same regressions using the main currencies that compose each industry-specific basket. We take as dependent variables the simple currency return (i.e., percentage change of the exchange rate) or 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, and one and two 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). Among other things, their model also predicts that equity market order flow driven by marketwide private information forecasts exchange rate returns. Our findings provide direct evidence con-
sistent 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 explain the home bias and return chasing of US investors, respectively. These papers implicitly hypothesize that there is 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 (2004) postulate that agents have aggregate private information that is used in both the money markets and the foreign exchange market. Albuquerque, Bauer and Schneider (2005) hypothesize that US 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 arbitraging where informed investors receive private signals about the only factor driving cash flows. Caballe and Krishnan (1994) build a general multi-agent, multi-security asset pricing framework with marketwide private information.

In spite of its prominent role in asset pricing theories, there has been little evidence presented in favor of the existence of marketwide private information. 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 with the estimation of non-structural models. Barclay et al. (1990) looking at the relation of index return volatility and trading volume conclude in favor of 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 US investors on eight developed country equity markets. Their measure is a statistical decomposition of flows that reflects covariation in unexpected net-flows.
of US 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, which they interpret as having prices that incorporate greater marketwide (private and public) information.

Our results complement those in Evans and Lyons (2002, 2004b) and Rime (2001), who link currency order flow with currency returns, and those in the domestic microstructure literature (e.g., Hasbrouck (1991)), who links stock order flow with own stock return. 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 et al. (2006) who study the reverse information-spillover of currency order flow into the equity market, finding mostly effects in volatilities.

In section 2 we develop a theoretical model of trading which allows us to estimate a measure of marketwide private information and to conduct our hypotheses tests. In section 3 we give details on the data used. Section 4 presents results on the estimation of marketwide private information and on tests of the main hypotheses. Section 5 concludes. The Appendix gives additional details on the sample selection, the currency exposure of the firms in our sample, and properties of the model estimates.

2 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 the component of observed order flow that is due to marketwide private information from all else using a structural model.

2.1 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 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 in EKOP and Easley and

---

2 Evans and Lyons (2004b) show that order flow in foreign currency markets can also forecast output growth, money growth and inflation.
O’Hara (1992), there is no explicit feedback from prices to order flow. Hence prices play no role in our analysis of order flow below and their presentation is omitted from the paper (see EKOP for details). We deviate from EKOP by allowing investors and the market maker to trade on $I > 1$ risky stocks indexed by $i = 1, \ldots, I$, and cash balances.

Prior to the start of any trading day informed investors may receive up to $I + 1$ bits of information. First, with probability $\theta$ there is an information event containing marketwide private information which is useful across all $I$ assets. Such 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. Second, regarding firm $i$, with probability $\alpha_i$ there is an information event containing firm-specific private information which is useful only for firm $i$. Such 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 any day for firm $i$ together with the trading activity that is generated on average at each node. There can be days with both marketwide private information and firm-specific private information, only the former or only the later, or none of the two. 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, consistent with the view that marketwide private information is generally composed of less precise information. This assumption is reflected in the absence of $\mu^{a}_i$ (see below) in lines 4 and 7 on the right-hand-side of Figure 1. Informed investors may therefore not act in accordance to their marketwide private information.

Consider for example a day with good marketwide private information. There are three possible outcomes. If there is no firm-specific information event for firm $i$, overall news on firm $i$ is good. If there is good firm-specific private information, then the two bits of information are reinforcing and overall news is good. Finally, if there is 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 $i$,
Figure 1: Tree diagram of the trading process for stock $i$.

Let $\mu_f^i$ be the average arrival rate of informed investors who trade based on firm-specific private information news and let $\mu_a^i$ be the average arrival rate of informed investors who trade based on marketwide private information news. The index $i$ on both $\mu_f^i$ and $\mu_a^i$ means that each firm is allowed different sensitivities to firm-specific and marketwide factors. Uninformed investors buy orders arrive at an average rate $\epsilon_b^i$ and sell orders arrive at an average rate $\epsilon_s^i$. These parameters are constant for the duration of the trading period and known to everyone, but the market maker does not know if he is trading against an informed or an uninformed investor.

With this notation, consider the node for firm $i$ obtained on a day with bad marketwide and firm-specific private information news at the top of Figure 1. The total volume of sell orders
has an average of $\mu_i^f + \mu_i^a + \varepsilon_i^b$ whereas only uninformed investors buy so the total volume of buy orders has an average of $\varepsilon_i^b$. 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 $\varepsilon_i^b$ and that of sell orders is $\varepsilon_i^s$.

The rest of the arrival rates at the end of each node is constructed in similar fashion.

### 2.2 The Likelihood Function

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

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

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

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

$$l_B (\{S_i, B_i\}) = \alpha_i (1 - \delta_i) e^{-\varepsilon_i^b} \frac{(\varepsilon_i^s)^{S_{in}}}{S_{in}!} \frac{e^{-\left(\mu_i^f + \mu_i^a + \varepsilon_i^b\right)}}{B_{in}!} + \alpha_i \delta_i e^{-\left(\mu_i^f + \mu_i^a + \varepsilon_i^b\right)} \frac{(\mu_i^f + \mu_i^a + \varepsilon_i^b)^{S_{in}}}{S_{in}!} \frac{e^{-\varepsilon_i^b}}{B_{in}!} + (1 - \alpha_i) e^{-\left(\mu_i^a + \varepsilon_i^b\right)} \frac{(\mu_i^a + \varepsilon_i^b)^{S_{in}}}{S_{in}!} \frac{e^{-\varepsilon_i^b}}{B_{in}!}.$$  

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

Finally, a day with no marketwide private information news occurs with probability $1 - \theta$, and in these days, the probability of observing the pair of sell and buy orders $\{S_i, B_i\}$ for firm
\[ l_0 \left( \{S_i, B_i\} \right) = \alpha_i (1 - \delta_i) e^{-\varepsilon_i^a} \left( \frac{e^{\varepsilon_i^b} S_{in}^{\mu_i^f + \varepsilon_i^b}}{B_{in}!} \right) + \alpha_i \delta_i e^{-\mu_i^f} \left( \frac{e^{\varepsilon_i^b} S_{in}^{\mu_i^f + \varepsilon_i^b}}{B_{in}!} \right) + (1 - \alpha_i) e^{-\varepsilon_i^a} \left( \frac{e^{\varepsilon_i^b} S_{in}^{\mu_i^f + \varepsilon_i^b}}{B_{in}!} \right). \]

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

We can now construct the likelihood function of the data. 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

\[ l \left( \left( S_{in}, B_{in} \right)_{i=1,\ldots,I} \right) = \theta (1 - \rho) \Pi_{i=1}^I l_G \left( \{S_{in}, B_{in}\} \right) + \theta \rho \Pi_{i=1}^I l_B \left( \{S_{in}, B_{in}\} \right) + (1 - \theta) \Pi_{i=1}^I l_0 \left( \{S_{in}, B_{in}\} \right). \]

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

\[ L \left( \left( S_{in}, B_{in} \right)_{n=1,\ldots,N; i=1,\ldots,I} \right) = \Pi_{n=1}^N l \left( \left( S_{in}, B_{in} \right)_{i=1,\ldots,I} \right). \]  

(1)

The likelihood function (1) is maximized to solve for \( \left( \alpha_i, \delta_i, \theta, \rho, \mu_i^a, \mu_i^f, e_i^a, e_i^b \right)_{i=1,\ldots,I} \), where we allow all parameters except for \( \theta \) and \( \rho \) to vary by firm. Because this problem does not admit a closed-form solution, we resort to numerical methods to estimate the model (see Section 4).

Allowing the parameters \( \mu_i^f, \mu_i^a, e_i^b \) and \( e_i^a \) 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 \( (\alpha_i, \delta_i) \), we need for every firm that there exist common time series patterns in its order flow. To see why this is the case note that conditional on marketwide private information the ‘daily’ likelihoods are trinomials of Poisson probability functions which are bilinear in \( \alpha_i \) and \( \delta_i \) (see Easley et al. (1997) and Vega (2006)). In the absence of informed trading, \( (e_i^a, e_i^b) \) 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, the role of \( (e_i^a, e_i^b) \) is still to capture average trading, while \( \mu_i^f \) measures the abnormal number of buy or sell orders that are firm-specific (Vega (2006)). With some qualifications addressed next, this is true too in our model.
In our model $\mu_i^a$ also 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 $\mu_i^a$ separately from $\mu_i^f$: when private marketwide and firm-specific news agree, informed trading is abnormally higher than when they do not. Finally, loosely speaking, our model uses the time series fluctuations in average (across firms) order flow to identify the common parameters $(\theta, \rho)$, with days characterized by common patterns across firms pushing estimates of $\theta$ up, and among these, days characterized by common high levels of sell (buy) orders pushing estimates of $\rho$ up (down).

In contrast to Easley et al. (1997) and Vega (2006), trades posted to each firm are not independent in the sense that informed investors have aggregate news useful to trade across all firms. This implies that the estimation of firm $i$’s parameters $\alpha_i$, $\delta_i$, $\mu_i^a$, $\mu_i^f$, $\varepsilon_i^a$ and $\varepsilon_i^b$ depends on the estimation of the other firms’ parameters as they are linked by the arrival of marketwide private information news. To see this consider solving a maximum likelihood problem as in Easley et al. (1997) and Vega (2006), who study every firm in isolation, when in fact the true model is one where there is marketwide private information. Biases in the estimation of $\alpha_i$, $\varepsilon_i^a$ and $\varepsilon_i^b$ can occur, for example, if days that are perceived by such optimization as days of no firm-specific private information news are truly days with marketwide private information. This is more likely to occur if the true $\mu_i^a$ is sufficiently small, leading to an upward bias in $\varepsilon_i^a$ and $\varepsilon_i^b$. If, in contrast, the true $\mu_i^a$ is close to the true $\mu_i^f$, a day with marketwide private information will be mistaken for a day with firm-specific private information, biasing the value of $\mu_i^f$ downwards (upwards) if $\mu_i^f > \mu_i^a$ ($<\)$ and affecting the inference of $\alpha_i$ and $\delta_i$ as well. It is easy to construct more such examples, but 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 non-linear 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, $\mu_i^a$. The model also allows some firms to be negatively affected by shocks to a marketwide factor while the majority of firms is positively affected by the same shock (or vice-versa), by capturing the marketwide shocks as firm-specific events for the firms negatively affected. However, this requires that such marketwide shocks be infrequent.

---

3 The optimization thus assigns to liquidity trading the values $\varepsilon_i^a + \mu_i^a$ and $\varepsilon_i^b + \mu_i^a$ for those days.
events, because 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 estimation error on \( MPI \). To minimize this bias we have designed sample selection criteria that identify industries where firms face significant exposure to currency risk and are affected in similar ways by that same risk. Section 3.1 and Appendices A and B contain the details.\(^4\)

### 2.3 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 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 either they have no firm-specific private information on firm \( i \) (with probability \( 1 - \alpha_i \)) or have good firm-specific private information (with probability \( \alpha_i (1 - \delta_i) \)):

\[
\text{Average Marketwide-Informed Buys} = \theta (1 - \rho) \sum_{i=1}^{I} [1 - \alpha_i + \alpha_i (1 - \delta_i)] \mu_i^a. \quad (2)
\]

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

\[
\text{Average Marketwide-Informed Sells} = \theta \rho \sum_{i=1}^{I} (1 - \alpha_i + \alpha_i \delta_i) \mu_i^a. \quad (3)
\]

Combining (2) and (3), the industry daily average (across \( I \)) order flow driven by marketwide private information news is given by:

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

\(^4\)There is an alternative approach of explicitly modeling this type of heterogeneity with the advantage of being potentially applicable to any industry. The drawbacks consist of increased computational cost and loss of degrees of freedom. Suppose there is a single source of marketwide news. 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 do 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, bringing us back to our approach.

10
We make use of the variable $MPI$ to capture 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 as measure of private information the probability of informed trading, which they call $PIN$. $PIN$ does not distinguish between good news days versus 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 if one wishes to forecast 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 cannot be used to forecast actual returns.

2.4 $MPI$ and Stock and Currency Returns

This subsection describes two hypotheses 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 where it was obtained. Therefore, the first hypothesis we test is:

**Hypothesis 1** $MPI$ forecasts the equity returns of the firms in the industry where it was obtained.

In light of Hasbrouck’s identification scheme, finding evidence in favor of Hypothesis 1 is a necessary condition to claiming that indeed ours is a measure of private information. Hypothesis 1 is also a test on 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

---

\(^5\)Evans and Lyons (2004b) document that foreign exchange order flow has predictive power for aggregate variables including the exchange rate, but do not establish the connection with the stock market as implied by their model and as we do here.
follows because investors trade based on their private information, but set prices only based on public information. After trading, each investors’ private information becomes public as it 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 non-informative liquidity trades making marketwide private information trades not directly observable. It is thus likely that information contained in order flow diffuses slowly through time and remains useful to forecast currency returns further into the future (e.g. see 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 over and above the information content of current \( MPI \). We thus obtain the following hypothesis:

**Hypothesis 2** Equity order flow driven by marketwide private information, \( MPI \), forecast changes in exchange rate returns.

Finding evidence consistent with Hypothesis 2 is also consistent with \( MPI \) being a measure of marketwide information, i.e., it contains information that is relevant in a variety of markets.

In testing hypothesis 2 we are not interested in all trades due to private information, but only those trades that derive 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 asset trading model constructed above that allows the identification of equity market trades driven by marketwide private information from trades due to firm-specific private information and liquidity trades.

### 3 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.
3.1 Industry and Firm Selection

The method developed above 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 (6-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 is obtained from the US International Trade Commission database and shipment data is from the Annual Survey of Manufactures of the US Census Bureau. We start with the top 30 industries 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). This guarantees that firms are treated consistently through the period from January 1993 to December 2003. We make use of international trade data as a source of information on 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 9 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 structural 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, 1996). Firms that fail to meet the liquidity requirement on any given month are excluded from the estimation in that month. We exclude foreign firms that also did not have a significant presence in terms of operations in the US with the purpose of identifying firms with qualitatively similar exposures to exchange rates. For example, this criterion led us to include the U.K. firm Doncasters PLC in the Aircraft

---

6In 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.
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 those industries with at least four firms. The requirement for having several firms in an industry is justified by: (i) the need to identify those parameters that are firm-independent, i.e., the probability of marketwide private information $\theta$ and the fraction of time marketwide news are bad news $\rho$; and (ii) because 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 we end up with 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, i.e., the industry. The reasons for ‘entry’ include beginning 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, ending of trading in the NYSE, or changing of main business activity causing the firm to drop from the NAICS in consideration. In very few instances we were 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.

3.2 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, labelling an order as ‘buy’ if the transaction price is above the spread midpoint and as ‘sell’ if the transaction price is below the spread midpoint and leaving unclassified transactions at the spread midpoint. For the unclassified transactions they used 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. Downticks are sells. Zero-upticks (zero price changes in which 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 which implies that
a comparison of the execution price to the quotes in effect at the time of the transaction is
inappropriate. They propose a 5-second rule for comparing execution prices to quotes reported
a minimum of 5 seconds before the transaction was reported. Existing evidence suggests that
the Lee and Ready algorithm is a good method to identifying the direction of trade from data
in the TAQ database.\textsuperscript{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 is taken from Datastream and is complemented with data from the
International Financial Statistics of the International Monetary Fund. We use \textit{beginning of
month quotes} of foreign currency (FC) per US dollar ($ or USD), denoted by $S_{FC/\$t}$. Currency
returns in month $t$, $cr_t$, are given by

$$cr_t \equiv \ln S_{FC/\$t+1} - \ln S_{FC/\$t}.$$ 

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

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

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 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 re-
turns using as weights the previous month’s fraction of industry exports going to each country.
At each month we allow at most 5 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 is customs value imports for consumption from the US Trade
Commission database.

\textsuperscript{7}Odders-White (2000) studies the performance of the Lee and Ready method using TORQ data and finds
that it correctly classifies 85 percent 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 substantial information about the original orders can still be inferred from
trades and quotes data. 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 identified trade direction to examine the accuracy of
several trade classification algorithms. The Lee and Ready algorithm showed the best results, correctly classifying
81.05 percent of the trades.

\[\]
3.3 The Industries under the Microscope

This subsection describes each of the 5 industries used in the paper, documenting their foreign currency exposure and giving 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 it is a wide practice of U.S. firms to invoice in USD (e.g. Goldberg and Tille (2004)), which we also document for some of our firms. This does not mean that such firms are hedged against currency fluctuations, because 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).

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

This industry is composed of firms primarily engaged in manufacturing oil and gas field machinery and equipment, such as oil and gas field drilling and production machinery and equipment, 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 is the industry for which we have most firms and therefore for which we expect to obtain the best results.

This is the only industry for which we use import-weights as opposed to export-weights in the construction of the currency baskets and in the identification of the most relevant currencies. The reason for this choice is that several of the export markets suffered, during the sample period, severe currency crises and a linear model is not a good model to predict such sporadic
In our sample, total imports represented 96% of total shipments on average. The US 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.

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 where 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, 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.

Aircraft Engine and Engine Parts Manufacturing: NAICS 336412. This industry is engaged in manufacturing aircraft engines and engine parts, developing and making prototypes of aircraft engines and engine parts, aircraft propulsion system conversion, 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

---

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

9 Since January 1st, 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.
the USA, except for Doncasters which is incorporated in the UK. We have included Doncasters in our sample, because it has a sizeable share of its operations in the US.

The export-weighted exchange rate was calculated 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 as reported in the Appendix.

**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 such items 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. The main currencies though are the British pound, the Japanese yen, and the Canadian dollar. During our sample exports as a fraction of total shipments averaged 57% for the industry.

**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 of 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). The year of 2003 was dropped as we could not estimate the model due to the large levels of buy and sell orders. Of these, 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 US, 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 weaker results for this industry relative to the others.

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%.

4 Results

This section gives details on the maximum likelihood estimation of (1) and the estimation of MPI. It lays out the regression specifications, exact hypotheses 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 present the equity and currency returns forecasting results.

4.1 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 $\hat{\Theta}_t \equiv (\hat{\alpha}_i, \hat{\delta}_i, \hat{\theta}, \hat{\rho}, \hat{\varepsilon}_s^i, \hat{\varepsilon}_b^i, \hat{\mu}_a^i, \hat{\mu}_f^i)_{i=1,\ldots,I}$. We stress that the likelihood maximization makes use of all the data to estimate each of firm $i$’s parameters, including data from other firms. It is therefore incorrect to suggest that only 22 observations from firm $i$ are being used to estimate firm $i$’s parameters $\left(\hat{\alpha}_i, \hat{\delta}_i, \hat{\varepsilon}_s^i, \hat{\varepsilon}_b^i, \hat{\mu}_a^i, \hat{\mu}_f^i\right)$. Intuitively, the estimation of firm $i$’s parameters is not independent of the estimation of the other parameters as they are linked by the arrival of marketwide private information (see subsection 2.2 for a more detailed argument).

The constrained optimization that we have to solve for is highly dependent on initial conditions, particularly on the parameters $\mu_f^i$ and $\mu_a^i$. To minimize the impact of initial conditions we implement a grid search on the parameters $\mu_f^i$ and $\mu_a^i$. Specifically, we set the initial values...
of these parameters to be a fraction (not necessarily the same) of the difference between the maximum daily number of buy orders in a month to the mean daily value of buy orders in that month. The grid allows these fractions to vary between 0.1 and 1 with increments of 0.1. We set the initial values of \( \varepsilon_s^i \) and \( \varepsilon_b^i \) to each month’s mean daily value of sell and buy orders, respectively (Vega (2006)). We set the initial values of the remaining parameters \( \theta \), \( \rho \), \( \alpha_i \) and \( \delta_i \) to 0.5 as we have no prior regarding these probabilities. The estimated parameters are those that yield the highest value for the likelihood function. This procedure is applied over all months and industries for consistency. While this results in a very time-consuming procedure, especially because gradient search methods were of no use here, it makes us 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 did not improve and the estimation slowed down considerably.

In order to evaluate the overall quality of fit of the model, we proceed by analyzing the model fit for every estimation, i.e., every month, and also jointly, across months. In the first case, for every month \( t \), the hypothesis we test is:

\[
H_0^i: \alpha_i = \delta_i = \mu_i^f = \mu_i^a = \theta = \rho = 0, \forall i.
\]  

(4)

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, \( LR_t = -2(\log L_{0,t} - \log L_t) \sim \chi^2_{(4 \times I_t + 2)} \), where \( \log L_t \) 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

\[
H_0: \bigcap_{t=1}^T H_0^t.
\]  

(5)

The joint test \( H_0 \) is the intersection of \( T \) sub-hypothesis. Therefore, the decision rule is to reject \( H_0 \) if \( H_0^t \) is rejected for at least one \( t \), given the significance levels \( \{\alpha_t\} \). However, to test (5) we need the joint distribution of the \( LR_t \) statistics from (4), which is unknown unless we restrict the stochastic relationship across sub-samples; consequently the size of the test cannot

\(^{10}\) Under the null, the restricted log-likelihood for month \( t \) is:

\[
\log L = -N \sum_{i=1}^I \left( \varepsilon_s^i + \varepsilon_b^i \right) + \sum_{n=1}^N \sum_{i=1}^I \left( S_n \log \varepsilon_s^i + B_n \log \varepsilon_b^i \right) - \sum_{n=1}^N \sum_{i=1}^I \log (S_n!B_n!).
\]

It is straightforward to show that the maximum likelihood estimates in the restricted model are: \( \varepsilon_s^i = \frac{1}{N} \sum_{n=1}^N S_n \), and \( \varepsilon_b^i = \frac{1}{N} \sum_{n=1}^N B_n \), from which we can construct the maximum log-likelihood, \( \log L_{0,t} \).
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 $\alpha_t$ of the individual tests so that the induced test has 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, $\chi_1^2 = \frac{T}{\alpha}(4 \times I_t + 2)$, is the critical value of the joint test. Finally, $H_0$ is rejected if $LR_t > \chi_1^2 = \frac{T}{\alpha}(4 \times I_t + 2)$.

Table 1 shows the results of tests 4 and 5 above. The results show that for almost every month we reject the null of the noise trading model. It is then only natural that 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 model with private information.

Estimating the model for each month and industry means that we need not worry about possible non-stationarity of information with respect to the parameters $\mu_f^t$, $\mu_a^t$, $\epsilon_s^t$, and $\epsilon_b^t$. However, the observed growth in trading activity over the years requires that we detrend our measures of $MPI$. Easley et al. (2002) account for this non-stationarity endogenously in their estimation. With many firms this is cumbersome and costly to implement. Hence, we use the Hodrick-Prescott filter on 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 also means 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 required values higher than the largest positive floating-point number in our personal computers. For this reason the sample length varies across industries (see Table 1).

### 4.2 $MPI$ and Factors in Equity Order Flow

Table 2 characterizes estimated $MPI$ by its correlation with other data and model variables. Table 2 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% significance or better for only 2 of the industries.

We ask whether a simple factor analysis on order flow provides a good measure of mar-
ketwide private information that can be used in place of the more complex measure we derive from a structural model (see also subsection 4.5). Denote by $PC_1$ the detrended first principal component from intra-industry firm-level order flow in each month. $PC_1$ 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 then, Table 2 shows that the correlation between $PC_1$ 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 2). The results in Table 2 suggest that $PC_1$ is not a good proxy for $MPI$ as the correlation between $MPI$ and $PC_1$ is low and statistically insignificant in most industries.

This raises the question of what explains the large covariation in trades implied by $PC_1$. The answer lies in common variation in liquidity trades ($\hat{\varepsilon}_b^i$ and sells $\hat{\varepsilon}_s^i$). A simple test is the correlation between estimated aggregate liquidity trades $LT_t = \sum_{i=1}^I (\hat{\varepsilon}_b^i - \hat{\varepsilon}_s^i)$ and $PC_1$. These correlations are large and positive, and statistically significant. Further, we construct the first principal component from estimated liquidity trades $(\hat{\varepsilon}_b^i - \hat{\varepsilon}_s^i)_t$ of the firms in an industry, denoted by $LT_1$. $LT_1$ is a better descriptor of common variation in liquidity than the sum of liquidity trades across firms, $LT$. We ask how important is $MPI$ relative to $LT_1$ as a source of commonality in trades, as given by $PC_1$? To answer this question, we conduct a variance decomposition of the explained variance of $PC_1$ between $MPI$ and $LT_1$ and assign the covariance term in equal parts to $MPI$ and $LT_1$ as 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 $PC_1$ for industries 336411, 336412, 336413, and 331419, respectively, with $LT_1$ accounting for the remaining fraction (untabulated). The exception is industry 333132 where $MPI$ explains 98% of the variation in $PC_1$, whereas $LT_1$ explains only 2%. Therefore, the results indicate that common variation in liquidity trades is the most important source of common variation in order flow in four out of five industries. This finding can shed light on the statistical exercise in Hasbrouck and Seppi (2001) documenting significant co-movement in order flows of the 30 stocks in the Dow Jones Industrial.

Table 2 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 trades due to private information. It turns out that the negative correlation originates in small sample estimation error together with the fact that the estimation is able to capture 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, if more of these are explained by trading due to private information, then less is attributed to liquidity trading. To illustrate this intuition suppose for simplicity that there are no trades due to firm-specific private information and that all marketwide private information is good news. For simplicity also, suppose that $\theta$ is known. Then the maximum likelihood estimate of $\varepsilon_i^s$ is $N^{-1} \sum_{n=1}^{N} S_{in}$, as nothing else explains why people sell. The buy orders in contrast can either come from $\mu_i^a$ or $\varepsilon_i^b$. As the model tries to capture the total number of buy orders $N^{-1} \sum_{n=1}^{N} B_{in}$, any estimation error in $\mu_i^a$ affects negatively estimates of $\varepsilon_i^b$, but not estimates of $\varepsilon_i^s$. Hence, in this case, estimates of $LT_i = \varepsilon_i^b - \varepsilon_i^s$ vary negatively with estimates of informed trading $MPI_i = \theta \mu_i^a$.

Importantly, we do not think that estimation error, in either $\mu_i^a$, $\varepsilon_i^b$ or $\varepsilon_i^s$, matters for the interpretation of other correlations that we emphasize in the paper. First, we expect it to 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, i.e., in months when $PC1$ is also high, if anything, we expect estimation error to increase the correlation between $MPI$ and $PC1$. Because we find that the correlation between estimated $LT$ and $PC1$ is higher than that between $MPI$ and $PC1$, we take our results to be 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, $FPI_t = \sum_{i=1}^{I} \hat{\alpha}_{i,t} \left( 1 - 2 \hat{\delta}_{i,t} \right) \hat{\mu}_{i,t}$. 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 4 industries, the model appears to be doing a good job decomposing order flow and confirms in a different way the goodness of fit results described in the previous section.

The last column of Table 2 gives the numbers for industry 331419. It is apparent, by looking across industries, that this industry’s $MPI$ stands out and contrasts with the properties of $MPI$ for the other industries. We believe that this is related to the fact that this industry uses two
firms almost all of the time to estimate \( MPI \) as the other two firms in the industry only sporadically meet the liquidity criterion (see subsection 3.3).

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 non-index stocks. Industry effects exist but are also small. Even though in our data \( MPI \) is weakly correlated or even uncorrelated with the first principal component, the evidence in Harford and Kaul might still be viewed as problematic to 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 belong to an index, only 1 out of 6 in NAICS 336412 belong 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 is estimated \( MPI \) correlated across industries. This 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 5 estimated values of \( MPI \) reveals that the first component explains 46\% of the total variation and that the first 2 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 there are at most two sources of marketwide private information affecting these 5 industries. However, in the absence of more economic structure in the model it is difficult to further evaluate this hypothesis, since even if only one factor is present, the correlation across \( MPI \)s can be low due to different industry sensitivities to the factor and due to noise in estimated \( MPI \).

4.3 Specification of the Forecasting Regressions and Hypotheses Tests

4.3.1 Forecasting Equity Returns

For each industry, we regress firm stock returns on lagged values of \( MPI \):

\[
RET_{i,t+j} = a_{i,0} + \sum_{l=1,\ldots,L} a_l MPI_{t-l} + u_{i,t+j},
\]

\(^{11}\)Indexes considered were: 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.
where 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. The inclusion of lags in this regression follows Hasbrouck (1991) 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 the ones we use to estimate $MPI$. Hypothesis 1 is that $MPI$ forecasts equity returns:

$$H_0 : a_1 = \ldots = a_L = 0,$$

against $H_A : a_l \neq 0$, 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 $MPI$: $\sum_{l=1}^{L} a_l > 0$.

As indicated in (6), we test the impact of $MPI$ on equity returns using panel data methods. Specifically, we assume fixed effects in the intercept and use the Within-Groups estimator. The generalized Hausman test procedure rejected the random effects specification in every case as well as the random coefficients specification. The alternative is to aggregate firm returns into an industry return (based on the firms we use), but the model, however, 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 and not the industry weights and of rendering more degrees of freedom.

### 4.3.2 Forecasting Currency Returns

For each industry, we report results for two regression specifications using either currency returns or excess currency returns. Let $\Delta CUR_{t+j}$ 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:

$$\Delta CUR_{t+j} = \alpha_0 + \sum_{t=1, \ldots, 10} \alpha_t MPI_{t-l+1} + u_{t+j}. \quad (7)$$

The regressions are conducted with contemporaneous ($j = 0$), one month ahead ($j = 1$) and two months ahead ($j = 2$) currency and excess currency returns. The hypothesis we test

---

12 By this we mean, whenever the AIC turned out to be of the same magnitude for two or more lag lengths. Choosing in accordance with the BIC meant, as to be expected, the choice of the more parsimonious specification. This also meant that using the BIC instead of the AIC in every situation would deliver specifications at odds with the Wald significance tests.
(Hypothesis 2) is that MPI explains currency movements:

$$H_0 : \alpha_1 = ... = \alpha_{10} = 0,$$

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

In specification (7), 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 that 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 was often at odds with the Wald significance test leading us to reject globally significant regressions in favor of more parsimonious but non-significant specifications.

We have also estimated a constrained version of (7) where $$\alpha_2 = ... = \alpha_{10} = 0$$. This specification presumes that private information diffuses quickly and is fully incorporated into prices in one month. Instead, the specification in (7) 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 gradual diffusion of information hypothesis is strongly favored by the data and report those results alone to conserve on space.

The use of a generated regressor in (6) and (7) means that we have to account for possible errors-in-variables. As shown in Pagan (1984) and in 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. Inference, however, may be affected by the sampling variation of the estimator used to generate the regressor, which would require that the standard errors be corrected for the first step estimation (Newey and McFadden (1994)).\textsuperscript{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 (7), 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 $$H_0 : \alpha_l = 0$$, $$l = 1, ..., L$$,

\textsuperscript{13} An alternative is to use bootstrapping. In this case, however, this is not feasible as it requires bootstrapping not only the second step estimation, but also simultaneously the first and time-consuming MLE step.
the uncorrected Newey-West estimator of the asymptotic variance of the OLS estimator of regression coefficients in (7) is consistent (likewise for model (6) under the null $H_0 : a_l = 0, l = 1, ..., L$). Note that we use robust standard errors because of the autocorrelation induced by overlapping observations (see Hansen and Hodrick (1980)).

4.4 Forecasting Results

We start with the results for equity returns in Table 3. Each cell of Table 3 has three numbers: the first is the sum of the estimated coefficients $\sum_{l=1}^{L} \hat{a}_l$, the second is the $p-$value on the significance of the sum of the coefficients and the third number is the $p-$value on the null hypothesis $H_0 : a_l = 0, l = 1, ..., L$. Our estimated information shock, $MPI$, successfully forecasts returns in 4 out of 5 industries. For industries 333132, 336411, 336412 and 336413, not only do we reject the null of no explanatory power of $MPI$, but we also get positive significant estimates for the cumulative response of equity returns.14 For the one month ahead stock return regressions in industries NAICS 336411 and 336412 we get globally significant regressions, but with a sum of coefficients associated with current and lagged $MPI$ statistically insignificant. On the negative side, we cannot forecast returns in industry 331419 with $MPI$. While we are not certain why this failure occurs, we note our discussion in subsection 4.2 on why this industry may have trouble in generating a reasonable estimate of 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 than that of the others (13.09 versus 4.1-7.2 for the other industries) and the highest kurtosis (11.25 versus 4.8-7 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 and 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, respecively (Easley and O’Hara (1992))

14 It has been shown that the significance of regressions with asset returns as dependent variables can be overstated (Stambaugh (1999)). This is more so if one regresses stock returns on a lagged stochastic regressor that depends on prices, such as the dividend yield, as the OLS estimator will have an upward finite-sample bias. As Stambaugh argues, this is because, by definition, such an explanatory variable will not be orthogonal to the disturbance term in the predictive regression at all leads and lags. We do not feel the need to correct for this potential bias as our explanatory variable, $MPI$, is a measure of trade volume that does not directly depend on stock prices. This is even truer for the currency return regressions as $MPI$ relates to a different asset market. Furthermore, there is controversy on the use of this correction (Lewellen (2004)).
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 to trade on.

Now turn to the forecasting of currency returns in Tables 3-7. The summary of our results is: MPI forecasts currency returns as well as excess currency returns quite well: (i) MPI forecasts excess currency-basket returns slightly better than simple currency-basket returns; (ii) MPI can forecast currency baskets but also, importantly, the main individual currencies that compose these baskets; (iii) lags of MPI appear to have significant forecasting ability consistent with a slow diffusion of information through time (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 4 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 subsection 4.3 above. (This structure is repeated in Tables 4-8.) Because of the large number of firms with apparent similar exposures in this industry, we expect that this is one of the industries where MPI is better estimated. 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, one month and two month ahead currency returns and excess returns on every currency, including the currency basket. The exception is the British pound. MPI forecasts returns on the Canadian dollar, the most important currency in the industry.

Table 5 gives the results for the Aircraft Manufacturing industry (NAICS 336411). MPI shows significant predictive power for most currencies and specifications. In fact, across all selected currencies, only for contemporaneous and one 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

---

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 called for given our criterion, and if outside these devaluation periods.
of 20\% and in most instances the regressions are globally significant at the 1\% significance level.

Table 6 displays the results for the Aircraft Engine Manufacturing industry (NAICS 336412). 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. For the French franc we can still forecast contemporaneous and one month ahead currency returns, and for the German mark only contemporaneous ones. Interestingly, contemporaneous MPI alone can forecast one month ahead returns of the German mark, i.e. for \( j = 1 \) (results available upon request). The fact that a distributed lag structure of MPI cannot may be due to either overfitting of the regression or to the loss of observations due to the inclusion of the 9 lags, but this is clearly the exception in our results.

Table 7 shows the results for the Aircraft Parts Manufacturing industry (NAICS 336413). MPI forecasts currency returns and excess currency returns for every currency at several horizons, except for the Japanese yen. MPI can also forecast simple currency returns one and two 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 missidentification of currency exposures. Instead, it could simply be a case where the impact of measurement error is more strongly felt (note that predictability in simple FRF returns is only marginally rejected).

Finally, Table 8 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. This is in spite of the difficulty in forecasting stock returns in this industry. In terms of individual currencies, it does well with the Swiss franc, the single most important currency in terms of exports, the French franc and the Japanese yen. On the downside, it fails to forecast the second most important currency, the British pound, and again the \( R^2 \)'s are considerably lower than in the other industries. The poor performance of this industry is consistent with our suspicion that the model cannot estimate MPI properly.

It is apparent from our results 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, it is possible that the marketwide information contained in MPI is revealed through public signals more quickly. 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)).

4.5 Robustness Checks

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

Table 2 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 two months ahead) currency returns or excess currency returns at the 10% significance level. These results imply that our measure of market wide private information can better extract private information than the simple averaging out of order flow.

Another robustness check is to run 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 co-movement in order flow across firms, including some derived from marketwide private information. In this robustness exercise we determine which measure does a better job at 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 9 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, whenever this was the case before, 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 that 1 p.p.) on average for industry 331419 and about 8% (or 1.2 p.p.) 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. As MPI’s forecasting ability does not seem to be significantly affected by the principal component,
we conclude that $PC1$ is capturing some other source of explanatory power.

What are the potential sources of the explanatory power of $PC1$ for industries 336413 and 331419? The evidence presented in subsection 4.2 above that $LT1$ is an important driver of $PC1$ suggests that we look at what explains correlated liquidity flow and why would such trades have a price impact. First, it is possible that measurement error leads some of 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 co-movement and hence of liquidity trades. Third, it has been suggested that correlated noise trading risk can generate long-run price effects (DeLong et al. (1990)). 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. The fact that in their model noise traders earn higher expected returns would then predict the ability of $PC1$ to forecast returns. Campbell and Kyle (1993) also have a model where noise trading is correlated with fundamentals. The separate roles of $MPI$ and $PC1$ and their joint properties represent 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) being captured by the lags of $MPI$. For this, we performed the same regressions for each currency on lags of $MPI$ and lagged 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 delivered zero lags. The analysis is only done for contemporaneous and one month ahead currency returns as the overlapping observations-induced autocorrelation causes an orthogonality violation for the regressions on two month ahead currency returns. 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. In fact, 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.

---

$16$ 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.
4.6 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 later requires that we be thorough 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 2 and subsection 4.5 above that the first principal component in industry order flow PC1 is not a good substitute for MPI. Can we do a better job if instead of using PC1 we use the first principal component in order flow due to firm-specific trading estimated 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:

\[
l(S_i, B_i) = \prod_{n=1}^{N} \left[ \alpha_i (1 - \delta_i) e^{-\varepsilon_i} \left( \frac{(\varepsilon_i^f S_{in})}{S_{in}} \right) e^{-\left(\mu_i^f + \varepsilon_i^b\right) \left(\frac{B_{in}}{B_{in}}\right)} + \alpha_i \delta_i e^{-\left(\mu_i^f + \varepsilon_i^b\right) \left(\frac{S_{in}}{S_{in}}\right) e^{-\left(\varepsilon_i^b\right) \left(\frac{B_{in}}{B_{in}}\right)}} + (1 - \alpha_i) e^{-\varepsilon_i} \left(\frac{(\varepsilon_i^b B_{in})}{B_{in}}\right) \right].
\]

This is a much easier and faster problem to solve. Once estimates for the parameters are obtained, compute a measure of trades due to firm-specific private information:

\[FPI_i = \hat{\alpha}_i \left(1 - 2\delta_i\right) \hat{\mu}_i^f.\]

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 \( MPI' \). \( MPI' \) 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 there is entry and exit of firms and some firms may be temporarily excluded from our sample due to low liquidity. We deal with this by estimating the first principal component in the \( FPI_i's \) over each different block of data.

To evaluate \( MPI' \), we first compute the percentage of explained variation in MPI, our original measure, that is explained by the new measure \( MPI' \), i.e., 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 (untabulated). Second, we repeat the exercises leading to Tables 3 through 8 using MPI’. The results (available upon request) are broadly consistent with MPI’ being at best a noisier measure of marketwide private information than 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.

5 Conclusion

This paper presents a model of equity trading with informed and uninformed investors, where informed investors act upon firm-specific private information and marketwide private information. The model is used to structurally identify and 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 as well as 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.

It is an interesting question whether there is any information that helps market makers (and econometricians) better understand the nature of the information that is about to hit the equity market. 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 trading on that information occurs first in such markets as they are more liquid, then by conditioning on currency order flow it is possible to improve the fit of the model.
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 6-digit NAICS industries that have the highest levels of exports in each of the years from 1997 to 2003. Data from exports is FAS value of domestic exports obtained from the US International Trade Commission database, available at http://dataweb.usitc.gov/. Our focus on manufacturing industries (i.e., NAICS 31-33) is without loss as manufacturing represents roughly 90% of all US exports. These data are merged with shipment data from the US 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 US Census Bureau has identified a significant understatement of shipments prior to 1997 using the NAICS system. Therefore, we base our industry selection using only data from 1997 to 2003.

We computed the ratio of domestic exports to total shipments for each of the years and ranked 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 in boldface letters.

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 goes 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 a loss of 11 industries. Column 10 of Table AI (“NAICS & SIC Comparable, 2003 Rank”) lists the 9 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 through the sample.

Table AII lists the 9 industries from Table AI which include the 5 industries we study (in boldface letters). 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 later 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 for estimation of marketwide private information. Note that 4 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 the number of firms is also large. Industry NAICS 334513 is dropped as it ranked last across all the remaining 6 industries in terms of volume of exports (with 20% less exports than the next to last industry NAICS 333132), leaving us with 5 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 US and/or the fact that the US 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 on space we report only on information regarding representative years.

Oil and Gas Field Machinery and Equipment Manufacturing: NAICS 333132. In 2003 Baker Hughes reports having entered foreign currency forwards to partially hedge exposure to currency fluctuations in such currencies 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. In its 2001 annual report, it indicates that approximately 27% of net assets are impacted by changes in foreign currencies relative to the USD. Cooper Cameron has production facilities located in the United Kingdom
and other European and Asian countries. The firm’s profitability is eroded when the USD weakens against the British pound, the euro and other currencies. Indeed, Cooper Cameron was negatively impacted during 2003 as a result of the weakening USD and may be further negatively impacted if the USD 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 USD 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 reported in various 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 also exposed to fluctuations in the cross-rate UKP/Euro whereas Grant Prideco is exposed to Venezuelan and Chinese currencies.

**Aircraft Manufacturing: NAICS 336411.** In their annual reports, the firms in this industry acknowledged exposure to the Japanese yen, Australian dollar, Canadian dollar, and several European currencies. For instance, Boeing’s 2003 annual report acknowledges foreign currency exposure because of suppliers and subcontractors located in Europe while most operations are in the United States, Canada and Australia. Even though 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 USD, Boeing’s operating exposure remains; its competitiveness is affected by USD movements. As with Boeing, Grumman Corp also hedges most of its foreign currency transactions exposure. However, they do not hedge translation exposure resulting from operations abroad.

**Aircraft Engine and Engine Parts Manufacturing: NAICS 336412.** Sequa reports primary foreign currency exposure to the British pound/USD rate and the British pound/euro rate from its U.K. firm Warwick International. UNC Inc has significant foreign operations but uses natural hedges and financial hedges to eliminate most of the transactions exposure it faces. However, it does not hedge any of its translation exposure from foreign currency net assets. United Technologies had at a 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/USD and the British pound/euro
In 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.

**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 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 exposure to fluctuations in foreign currencies for transactions denominated primarily in the British pound, French franc, and Singapore dollar and hedges most of it. Sundstrand does not hedge any of 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.

**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 as well. For example, in 2001 Tremont’s annual report indicated that earnings were primarily affected by fluctuations in the value of the USD relative to the euro, the Canadian dollar, the Norwegian kroner, and the British pound. In 2001 TIMET, Tremont’s main subsidiary, had approximately 40% of its sales revenue originated in Europe, of which 60% was denominated in currencies other than the USD, 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/USD 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 $\mu_a^i$ (or $\mu_f^i$) are negatively correlated with estimates of $\epsilon^b_i$ and $\epsilon^s_i$, therefore inducing a correlation between informed trading and liquidity trading. Intuitively, as the model tries to match the total number of buy, or sell orders, if more of these are explained by marketwide driven private information trading, then less will have to be 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 $\hat{LT}_i$ and $\hat{MPI}_i$ for firm $i$ is
\[
\frac{1}{T} \sum_t \left( \hat{LT}_{it} - \bar{LT}_i \right) \times \left( \hat{MPI}_{it} - \bar{MPI}_i \right),
\]
where $\bar{LT}_i = \frac{1}{T} \sum_t \hat{LT}_{it}$ and $\bar{MPI}_i = \frac{1}{T} \sum_t \hat{MPI}_{it}$. With random samples of buy and sell orders drawn over time and with the assumption of constant distributional parameters, $\bar{LT}_i \rightarrow LT_i$ and $\bar{MPI}_i \rightarrow MPI_i$ 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 $\hat{LT}_{it}$ and $\hat{MPI}_{it}$ co-move. Notice that we shifted the analysis from the covariance between industry liquidity trading and marketwide private information trades 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, i.e., $\alpha_i = 0$ 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 instead we assume that only firm-specific private information exists.\(^\text{17}\) Finally, we assume that the estimation applies to the parameters $\mu_a^i$, $\epsilon^b_i$ and $\epsilon^s_i$ for all $i$, and that $\theta$ and $\rho$ are known.

Under these assumptions, equations (1)-(3) in the main text become:
\[
\theta \left( 1 - \rho \right) \prod_{i=1}^I \prod_{\{S_i, B_i\}} \left( e^{-\theta \left( \frac{\epsilon^s_i}{\epsilon^s_i} \right) S_{in} \frac{\epsilon^b_i}{\epsilon^b_i} B_{in} \left( \frac{\mu_a^i}{\epsilon^s_i} + 1 \right) B_{in} } \right),
\]
\[
\theta \rho \prod_{i=1}^I \prod_{\{S_i, B_i\}} \left( e^{-\rho \left( \frac{\epsilon^s_i}{\epsilon^s_i} \right) S_{in} \frac{\epsilon^b_i}{\epsilon^b_i} B_{in} \left( \frac{\mu_a^i}{\epsilon^s_i} + 1 \right) S_{in} } \right).
\]
\(^\text{17}\)We have not been able to obtain conditions that apply when both $\mu_a^i$ and $\mu_f^i$ co-exist. By continuity, our argument applies as either $\mu_a^i$ or $\mu_f^i$ approach zero. However, in simulations, it is possible to construct examples where the correlation between estimated $MPI$ and estimated $LT$ are positive.
and
\[(1 - \theta) \prod_{i=1}^{l} (\{S_i, B_i\}) = (1 - \theta) \prod_{i=1}^{l} \left[ e^{-\xi_i^b} \frac{(\xi_i^s)_{S_i}}{S_i!} e^{-\xi_i^b} \frac{(\xi_i^b)_{B_i}}{B_i!} \right]. \tag{C3} \]

The log-likelihood of observing \( I \times N \) buy and sell orders \( \{B_i, S_i\} \), \( \log L \left( \{S_i, B_i\}_{i=1}^{n} \right) \), is
\[
\sum_{n=1}^{N} \sum_{i=1}^{I} \left[ -\left( \xi_i^b + \xi_i^s \right) + S_i \log \xi_i^s + B_i \log \xi_i^b - \log \left( S_i! B_i! \right) \right] + \sum_{n=1}^{N} \log D_n, \tag{C4} \]
where
\[
D_n \equiv \theta (1 - \rho) \prod_{i=1}^{I} e^{-\mu_i^a} \left( \frac{\mu_i^a}{\xi_i^b} + 1 \right)^{B_i} + \theta \rho \prod_{i=1}^{I} e^{-\mu_i^a} \left( \frac{\mu_i^a}{\xi_i^b} + 1 \right)^{S_i} + 1 - \theta. \tag{C5} \]

Denote the average number of buy orders for firm \( i \) by \( \tilde{B}_i = \frac{1}{N} \sum_{n=1}^{N} B_i \) and the average number of sell orders by \( \tilde{S}_i = \frac{1}{N} \sum_{n=1}^{N} S_i \). The first order conditions for \( \xi_i^b \) and \( \xi_i^s \), evaluated at the MLE can be expressed as:
\[
\xi_i^b = \frac{1}{N} \sum_{n=1}^{N} B_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \tilde{\xi}_i^b} D_n^{-1} \theta (1 - \rho) \prod_{j=1}^{I} e^{-\hat{\mu}_j^a} \left( \frac{\hat{\mu}_j^a}{\tilde{\xi}_j^b} + 1 \right)^{B_{jn}}, \tag{C6} \]
\[
\xi_i^s = \frac{1}{N} \sum_{n=1}^{N} S_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \tilde{\xi}_i^s} D_n^{-1} \theta \rho \prod_{j=1}^{I} e^{-\hat{\mu}_j^a} \left( \frac{\hat{\mu}_j^a}{\tilde{\xi}_j^s} + 1 \right)^{S_{jn}}. \tag{C7} \]

We omit the first order conditions for the \( \mu_i^a \) as these are uninformative.

Under the additional restriction of no marketwide private information, i.e. \( \mu_i^a = 0 \), then
\[
\xi_i^b = \bar{B}_i, \quad \xi_i^s = \bar{S}_i, \tag{C8} \]
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 \( \mu_i^a \) assume that, for all \( j \) and all \( n, \)
\[
\left( \frac{\hat{\mu}_j^a}{\tilde{\xi}_j^b} + 1 \right)^{B_{jn}} \approx e^{\tilde{\xi}_j^b}, \quad \left( \frac{\hat{\mu}_j^a}{\tilde{\xi}_j^s} + 1 \right)^{S_{jn}} \approx e^{\tilde{\xi}_j^s}. \tag{C9} \]
These would be exact if \( \xi_j^b = B_{jn} \to \infty \) and \( \xi_j^s = S_{jn} \to \infty \). They are a reasonable approximation when liquidity trading is large relative to all other sources of trading and given the fact that large buy and sell orders push up the estimated parameters \( \xi_j^b \) and \( \xi_j^s \) (see (C8)). With this approximation the first order conditions simplify to
\[
\xi_i^b \simeq \tilde{B}_i - \tilde{B}_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \tilde{\xi}_i^b} \theta (1 - \rho), \tag{C10} \]
\[
\xi_i^s \simeq \tilde{S}_i - \tilde{S}_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \tilde{\xi}_i^s} \theta \rho. \tag{C11} \]
We next show that estimated $LT_i = \varepsilon_i^b - \varepsilon_i^s$ varies negatively with informed trading, which when $\alpha_i = 0$ becomes $MPI_i = (1 - 2\rho) \mu_i^a$. Consider two extreme cases for parameter $\rho$.

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

\[
\frac{\partial \hat{LT}_i}{\partial MPI_i} = \frac{\partial \hat{\varepsilon}_i^b}{\partial \hat{\mu}_i^a} = -\frac{\hat{\varepsilon}_i^b - \bar{B}_i (1 - \theta)}{\hat{\mu}_i^a - \bar{B}_i + 2\hat{\varepsilon}_i^b} < 0,
\]

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

\[
\hat{\varepsilon}_i^b \approx \bar{B}_i - \bar{B}_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \varepsilon_i^b} \theta > \bar{B}_i - \bar{B}_i \theta > 0.
\]

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 $\varepsilon_i^b = \bar{B}_i$, $MPI_i = -\mu_i^a$, and

\[
\frac{\partial \hat{LT}_i}{\partial MPI_i} = \frac{\partial \hat{\varepsilon}_i^s}{\partial \hat{\mu}_i^a} = -\frac{\hat{\varepsilon}_i^s - \bar{S}_i (1 - \theta)}{\hat{\mu}_i^a - \bar{S}_i + 2\hat{\varepsilon}_i^s} < 0,
\]

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

\[
\hat{\varepsilon}_i^s \approx \bar{S}_i - \bar{S}_i \frac{\hat{\mu}_i^a}{\hat{\mu}_i^a + \varepsilon_i^s} \theta > \bar{S}_i - \bar{S}_i \theta > 0.
\]

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

General case: \( \rho \in [0,1] \). Using (1)-(1) it is easy to show that

\[
\frac{\partial \hat{LT}_i}{\partial \hat{\mu}_i^a} = -\frac{\hat{\varepsilon}_b - \hat{B}_i (1 - \theta (1 - \rho))}{\hat{\mu}_i^a - B_i + 2\hat{\varepsilon}_b^a} + \frac{\hat{\varepsilon}_s - \hat{S}_i (1 - \theta \rho)}{\hat{\mu}_i^a - S_i + 2\hat{\varepsilon}_s^a}.
\]

(C16)

While it is difficult to sign this derivative in general, for low values of \( \mu_i^a \) we can go a long way. With \( \mu_i^a = 0 \), then \( \hat{\varepsilon}_s = \hat{S}_i \) and \( \hat{\varepsilon}_b = \hat{B}_i \) (see (C8)), implying that:

\[
\frac{\partial \hat{LT}_i}{\partial \hat{\mu}_i^a} \bigg|_{\mu_i^a = 0} = -\theta (1 - 2\rho).
\]

(C17)

Given that \( \frac{\partial MPI_i}{\partial \mu_i^a} = 1 - 2\rho \) we have

\[
\frac{\partial \hat{LT}_i}{\partial MPI_i} \bigg|_{\mu_i^a = 0} = -\theta < 0.
\]

(C18)

Therefore, a negative correlation obtains.
References


The table shows the fraction of months in which the noise trading model (hypothesis \( H_0: \alpha_i = \delta_i = \mu_i^f = \mu_i^a = \theta = \rho = 0, \forall i \)) 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 \( H_0: \bigcap_{t=1}^{T} H_0 \) are LR and \( \chi^2_{1-\alpha/T}(4 \times I_t + 2) \), respectively, 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.

<table>
<thead>
<tr>
<th>NAICS</th>
<th>NAICS</th>
<th>NAICS</th>
<th>NAICS</th>
<th>NAICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>333132</td>
<td>336411</td>
<td>336412</td>
<td>336413</td>
<td>331419</td>
</tr>
</tbody>
</table>

**Individual Goodness of Fit Tests**
Fraction of rejections of \( H_0^t \) in %

| 97.3 | 98.8 | 100.0 | 99.1 | 100.0 |

**Joint Goodness of Fit Test**
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 |
Table 2
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.

<table>
<thead>
<tr>
<th>Correlation Between</th>
<th>NAICS 333132</th>
<th>NAICS 336411</th>
<th>NAICS 336412</th>
<th>NAICS 336413</th>
<th>NAICS 331419</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPI and TOF</td>
<td>0.3258</td>
<td>0.1602</td>
<td>0.1578</td>
<td>0.1783</td>
<td>0.3180</td>
</tr>
<tr>
<td></td>
<td>0.0785</td>
<td>0.2785</td>
<td>0.3765</td>
<td>0.1279</td>
<td>0.0120</td>
</tr>
<tr>
<td>MPI and PC1</td>
<td>0.2911</td>
<td>0.0447</td>
<td>0.1299</td>
<td>0.1244</td>
<td>0.3016</td>
</tr>
<tr>
<td></td>
<td>0.0817</td>
<td>0.7791</td>
<td>0.3936</td>
<td>0.1950</td>
<td>0.0200</td>
</tr>
<tr>
<td>MPI and LT</td>
<td>-0.6585</td>
<td>-0.6678</td>
<td>-0.7173</td>
<td>-0.5281</td>
<td>0.0917</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.4498</td>
</tr>
<tr>
<td>MPI + LT and TOF</td>
<td>0.6401</td>
<td>0.7209</td>
<td>0.5002</td>
<td>0.6612</td>
<td>0.3517</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0036</td>
</tr>
<tr>
<td>MPI+LT and PC1</td>
<td>0.4889</td>
<td>0.7220</td>
<td>0.5197</td>
<td>0.6265</td>
<td>0.3853</td>
</tr>
<tr>
<td></td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0012</td>
</tr>
<tr>
<td>LT and TOF</td>
<td>0.3426</td>
<td>0.5039</td>
<td>0.2805</td>
<td>0.5170</td>
<td>0.2643</td>
</tr>
<tr>
<td></td>
<td>0.0272</td>
<td>0.0007</td>
<td>0.0340</td>
<td>0.0000</td>
<td>0.0034</td>
</tr>
<tr>
<td>LT and PC1</td>
<td>0.2418</td>
<td>0.5516</td>
<td>0.3091</td>
<td>0.3581</td>
<td>0.3080</td>
</tr>
<tr>
<td></td>
<td>0.0958</td>
<td>0.0001</td>
<td>0.0147</td>
<td>0.0321</td>
<td>0.0103</td>
</tr>
<tr>
<td>FPI and TOF</td>
<td>0.4214</td>
<td>0.2408</td>
<td>0.0678</td>
<td>0.1956</td>
<td>0.2259</td>
</tr>
<tr>
<td></td>
<td>0.0052</td>
<td>0.1755</td>
<td>0.6311</td>
<td>0.0899</td>
<td>0.0589</td>
</tr>
<tr>
<td>FPI and MPI</td>
<td>0.6874</td>
<td>0.6645</td>
<td>0.7373</td>
<td>0.5133</td>
<td>0.2400</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0833</td>
</tr>
<tr>
<td>FPI and LT</td>
<td>-0.6334</td>
<td>-0.6344</td>
<td>-0.8935</td>
<td>-0.6861</td>
<td>0.0676</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0007</td>
<td>0.5491</td>
</tr>
<tr>
<td>FPI and PC1</td>
<td>0.4732</td>
<td>0.1511</td>
<td>0.0363</td>
<td>0.1770</td>
<td>0.2287</td>
</tr>
<tr>
<td></td>
<td>0.0038</td>
<td>0.4533</td>
<td>0.7985</td>
<td>0.0393</td>
<td>0.0517</td>
</tr>
<tr>
<td>MPI+FPI+LT and TOF</td>
<td>0.9767</td>
<td>0.9688</td>
<td>0.7437</td>
<td>0.9824</td>
<td>0.3751</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0042</td>
</tr>
<tr>
<td>MPI+FPI+LT and PC1</td>
<td>0.8708</td>
<td>0.9024</td>
<td>0.7287</td>
<td>0.9364</td>
<td>0.4080</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0015</td>
</tr>
<tr>
<td>PC1 and TOF</td>
<td>0.8736</td>
<td>0.8757</td>
<td>0.9752</td>
<td>0.9474</td>
<td>0.9632</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>114</td>
<td>84</td>
<td>96</td>
<td>121</td>
<td>120</td>
</tr>
</tbody>
</table>
**Table 3**

**Stock Returns and Order Flow Driven by Market-wide Private Information.**

The table shows statistics obtained from the regression

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

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 market-wide 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 \( \sum_{l=1,...,L} a_l \), the second number gives the \( p \)-value on the hypothesis that the sum of 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 obtained with robust standard errors.

<table>
<thead>
<tr>
<th>Coefficients and ( p )-values</th>
<th>j</th>
<th>NAICS</th>
<th>NAICS</th>
<th>NAICS</th>
<th>NAICS</th>
<th>NAICS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A</strong></td>
<td></td>
<td>333132</td>
<td>336411</td>
<td>336412</td>
<td>336413</td>
<td>331419</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.0007</td>
<td>0.0004</td>
<td>-0.0000</td>
<td>0.0027</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0140</td>
<td>0.6660</td>
<td>0.9900</td>
<td>0.0620</td>
<td>0.3754</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0024</td>
<td>0.0004</td>
<td>0.3754</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.0007</td>
<td>0.0003</td>
<td>0.0020</td>
<td>0.0043</td>
<td>-0.0005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0080</td>
<td>0.0718</td>
<td>0.0110</td>
<td>0.0060</td>
<td>0.5531</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0718</td>
<td>0.0693</td>
<td>0.0002</td>
<td>0.5531</td>
</tr>
<tr>
<td><strong>Panel B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60-day</td>
<td></td>
<td>0.0014</td>
<td>0.0006</td>
<td>0.0269</td>
<td>0.0073</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0320</td>
<td>0.0020</td>
<td>0.0010</td>
<td>0.8629</td>
</tr>
<tr>
<td>90-day</td>
<td></td>
<td>0.0000</td>
<td>0.0672</td>
<td>0.0934</td>
<td>0.0000</td>
<td>0.8629</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0022</td>
<td>0.0010</td>
<td>0.0030</td>
<td>0.0110</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0030</td>
<td>0.0040</td>
<td>0.0000</td>
<td>0.5478</td>
</tr>
<tr>
<td>120-day</td>
<td></td>
<td>0.0000</td>
<td>0.0100</td>
<td>0.0118</td>
<td>0.0000</td>
<td>0.5478</td>
</tr>
<tr>
<td>90-day</td>
<td></td>
<td>0.0030</td>
<td>0.0008</td>
<td>0.0023</td>
<td>0.0152</td>
<td>0.0013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0024</td>
<td>0.0060</td>
<td>0.0000</td>
<td>0.3334</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0024</td>
<td>0.0150</td>
<td>0.0000</td>
<td>0.3334</td>
</tr>
</tbody>
</table>
Table 4
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

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

In panel A, \( \Delta CUR_{t+j} \) 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, \( H_0: \alpha_l = 0, l > 0 \). 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).

<table>
<thead>
<tr>
<th>Currency</th>
<th>CAD</th>
<th>DEM</th>
<th>FRF</th>
<th>GBP</th>
<th>ITL</th>
<th>NLG</th>
<th>NOK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basket</td>
<td>26.09%</td>
<td>5.33%</td>
<td>6.87%</td>
<td>17.63%</td>
<td>5.74%</td>
<td>7.79%</td>
<td>3.94%</td>
</tr>
</tbody>
</table>

**Panel A: Currency Returns**

\( j = 0 \)

<table>
<thead>
<tr>
<th></th>
<th>0.1200</th>
<th>0.1272</th>
<th>0.1417</th>
<th>0.1451</th>
<th>0.0575</th>
<th>0.1822</th>
<th>0.1445</th>
<th>0.1370</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.6866</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

\( j = 1 \)

<table>
<thead>
<tr>
<th></th>
<th>0.1644</th>
<th>0.1796</th>
<th>0.1585</th>
<th>0.1653</th>
<th>0.0452</th>
<th>0.1928</th>
<th>0.1620</th>
<th>0.1433</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.7097</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

\( j = 2 \)

<table>
<thead>
<tr>
<th></th>
<th>0.1184</th>
<th>0.1621</th>
<th>0.1371</th>
<th>0.1397</th>
<th>0.0490</th>
<th>0.1732</th>
<th>0.1394</th>
<th>0.1463</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0015</td>
<td>0.0027</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.1059</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

**Panel B: Excess Currency Returns**

\( j = 0 \)

<table>
<thead>
<tr>
<th></th>
<th>0.1107</th>
<th>0.1259</th>
<th>0.1397</th>
<th>0.1433</th>
<th>0.0584</th>
<th>0.1821</th>
<th>0.1428</th>
<th>0.1309</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0009</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.6313</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

\( j = 1 \)

<table>
<thead>
<tr>
<th></th>
<th>0.1672</th>
<th>0.1800</th>
<th>0.1607</th>
<th>0.1674</th>
<th>0.0450</th>
<th>0.1952</th>
<th>0.1649</th>
<th>0.1384</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0005</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.6704</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

\( j = 2 \)

<table>
<thead>
<tr>
<th></th>
<th>0.1288</th>
<th>0.1641</th>
<th>0.1419</th>
<th>0.1442</th>
<th>0.0497</th>
<th>0.1790</th>
<th>0.1252</th>
<th>0.1447</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0005</td>
<td>0.0018</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0930</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Table 5
Exchange Rates and Order Flow Driven by Market-wide Private Information in the Aircraft Manufacturing Industry, NAICS 336411

The table shows statistics obtained from the regression
\[ \Delta CUR_{t+j} = \alpha_0 + \sum_{l=1,...,10} \alpha_l MPI_{t-l+1} + u_{t+j}. \]

In panel A, \( \Delta CUR_{t+j} \) 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, \( H_0: \alpha_l = 0, l > 0 \).

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).

<table>
<thead>
<tr>
<th></th>
<th>Currency</th>
<th>CAD</th>
<th>DEM</th>
<th>FRF</th>
<th>GBP</th>
<th>JPY</th>
<th>NLG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basket</td>
<td></td>
<td>2.16%</td>
<td>4.57%</td>
<td>2.57%</td>
<td>9.31%</td>
<td>8.23%</td>
<td>3.79%</td>
</tr>
</tbody>
</table>

Panel A: Currency Returns

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>j = 0</td>
<td>0.0908</td>
<td>0.2222</td>
<td>0.1157</td>
<td>0.1117</td>
<td>0.0505</td>
<td>0.2535</td>
</tr>
<tr>
<td></td>
<td>0.4332</td>
<td>0.0000</td>
<td>0.0365</td>
<td>0.0505</td>
<td>0.4616</td>
<td>0.0000</td>
</tr>
<tr>
<td>j = 1</td>
<td>0.1198</td>
<td>0.2059</td>
<td>0.1413</td>
<td>0.1493</td>
<td>0.0605</td>
<td>0.2682</td>
</tr>
<tr>
<td></td>
<td>0.4562</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0010</td>
<td>0.3705</td>
<td>0.0000</td>
</tr>
<tr>
<td>j = 2</td>
<td>0.1207</td>
<td>0.2086</td>
<td>0.1486</td>
<td>0.1501</td>
<td>0.1790</td>
<td>0.2138</td>
</tr>
<tr>
<td></td>
<td>0.6177</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0461</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Panel B: Excess Currency Returns

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>j = 0</td>
<td>0.1059</td>
<td>0.2350</td>
<td>0.1149</td>
<td>0.1084</td>
<td>0.0560</td>
<td>0.2551</td>
</tr>
<tr>
<td></td>
<td>0.0200</td>
<td>0.0000</td>
<td>0.0415</td>
<td>0.0847</td>
<td>0.3319</td>
<td>0.0000</td>
</tr>
<tr>
<td>j = 1</td>
<td>0.1103</td>
<td>0.2180</td>
<td>0.1414</td>
<td>0.1464</td>
<td>0.0630</td>
<td>0.2703</td>
</tr>
<tr>
<td></td>
<td>0.0197</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0007</td>
<td>0.2948</td>
<td>0.0000</td>
</tr>
<tr>
<td>j = 2</td>
<td>0.0735</td>
<td>0.2153</td>
<td>0.1499</td>
<td>0.1502</td>
<td>0.1837</td>
<td>0.2150</td>
</tr>
<tr>
<td></td>
<td>0.0072</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0321</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Table 6
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

\[
\Delta CUR_{t+j} = \alpha_0 + \sum_{l=1}^{10} \alpha_l MPI_{t-l+1} + u_{t+j}.
\]

In panel A, \(\Delta CUR_{t+j}\) 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, \(H_0: \alpha_l = 0, l > 0\). 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).

<table>
<thead>
<tr>
<th>Currency</th>
<th>CAD</th>
<th>DEM</th>
<th>FRF</th>
<th>GBP</th>
<th>JPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basket</td>
<td>11.84%</td>
<td>8.44%</td>
<td>22.89%</td>
<td>14.24%</td>
<td>6.13%</td>
</tr>
</tbody>
</table>

**Panel A: Currency Returns**

<table>
<thead>
<tr>
<th></th>
<th>(j = 0)</th>
<th>(j = 1)</th>
<th>(j = 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAD</td>
<td>0.1482</td>
<td>0.1227</td>
<td>0.1335</td>
</tr>
<tr>
<td>DEM</td>
<td>0.0447</td>
<td>0.0482</td>
<td>0.0932</td>
</tr>
<tr>
<td>FRF</td>
<td>0.1267</td>
<td>0.1114</td>
<td>0.1203</td>
</tr>
<tr>
<td>GBP</td>
<td>0.1337</td>
<td>0.1186</td>
<td>0.1259</td>
</tr>
<tr>
<td>JPY</td>
<td>0.1095</td>
<td>0.0999</td>
<td>0.1293</td>
</tr>
<tr>
<td>(\alpha_0)</td>
<td>0.0010</td>
<td>0.0137</td>
<td>0.0188</td>
</tr>
<tr>
<td>(\alpha_1)</td>
<td>0.0471</td>
<td>0.0114</td>
<td>0.0630</td>
</tr>
<tr>
<td>(\alpha_2)</td>
<td>0.0137</td>
<td>0.0663</td>
<td>0.0684</td>
</tr>
<tr>
<td>(\alpha_3)</td>
<td>0.0181</td>
<td>0.0043</td>
<td>0.0005</td>
</tr>
<tr>
<td>(\alpha_4)</td>
<td>0.0517</td>
<td>0.0023</td>
<td>0.0000</td>
</tr>
<tr>
<td>(\alpha_5)</td>
<td>0.0012</td>
<td>0.1431</td>
<td>0.0459</td>
</tr>
</tbody>
</table>

**Panel B: Excess Currency Returns**

<table>
<thead>
<tr>
<th></th>
<th>(j = 0)</th>
<th>(j = 1)</th>
<th>(j = 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAD</td>
<td>0.1528</td>
<td>0.1256</td>
<td>0.1422</td>
</tr>
<tr>
<td>DEM</td>
<td>0.0444</td>
<td>0.0504</td>
<td>0.0933</td>
</tr>
<tr>
<td>FRF</td>
<td>0.1276</td>
<td>0.1119</td>
<td>0.1230</td>
</tr>
<tr>
<td>GBP</td>
<td>0.1342</td>
<td>0.1184</td>
<td>0.1278</td>
</tr>
<tr>
<td>JPY</td>
<td>0.1107</td>
<td>0.0998</td>
<td>0.1307</td>
</tr>
<tr>
<td>(\alpha_0)</td>
<td>0.0006</td>
<td>0.0045</td>
<td>0.0071</td>
</tr>
<tr>
<td>(\alpha_1)</td>
<td>0.8749</td>
<td>0.7891</td>
<td>0.5189</td>
</tr>
<tr>
<td>(\alpha_2)</td>
<td>0.0183</td>
<td>0.1431</td>
<td>0.2154</td>
</tr>
<tr>
<td>(\alpha_3)</td>
<td>0.0217</td>
<td>0.0769</td>
<td>0.2188</td>
</tr>
<tr>
<td>(\alpha_4)</td>
<td>0.0524</td>
<td>0.0053</td>
<td>0.0000</td>
</tr>
<tr>
<td>(\alpha_5)</td>
<td>0.0009</td>
<td>0.0022</td>
<td>0.0372</td>
</tr>
</tbody>
</table>
Table 7
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

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

In panel A, \( \Delta CUR_{t+j} \) 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 auto-correlation and heteroskedasticity-adjusted \( p \)-value of a Wald test on the null hypothesis of joint significance, \( H_0: \alpha_l = 0, l > 0 \). 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).

<table>
<thead>
<tr>
<th>Currency</th>
<th>CAD</th>
<th>DEM</th>
<th>FRF</th>
<th>GBP</th>
<th>JPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basket</td>
<td>9.38%</td>
<td>5.31%</td>
<td>6.38%</td>
<td>14.24%</td>
<td>14.01%</td>
</tr>
</tbody>
</table>

**Panel A: Currency Returns**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>0.0495</th>
<th>0.1045</th>
<th>0.0651</th>
<th>0.0694</th>
<th>0.1175</th>
<th>0.0329</th>
</tr>
</thead>
<tbody>
<tr>
<td>j = 0</td>
<td>0.0534</td>
<td>0.0781</td>
<td>0.0941</td>
<td>0.0606</td>
<td>0.0632</td>
<td>0.1033</td>
<td>0.0258</td>
</tr>
<tr>
<td></td>
<td>0.0076</td>
<td>0.0648</td>
<td>0.2071</td>
<td>0.1103</td>
<td>0.0006</td>
<td>0.8800</td>
<td></td>
</tr>
<tr>
<td>j = 2</td>
<td>0.0084</td>
<td>0.0706</td>
<td>0.0006</td>
<td>0.0042</td>
<td>0.0000</td>
<td>0.6035</td>
<td></td>
</tr>
</tbody>
</table>

**Panel B: Excess Currency Returns**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>0.0522</th>
<th>0.1054</th>
<th>0.0783</th>
<th>0.0727</th>
<th>0.1171</th>
<th>0.0342</th>
</tr>
</thead>
<tbody>
<tr>
<td>j = 0</td>
<td>0.0619</td>
<td>0.0960</td>
<td>0.0717</td>
<td>0.0676</td>
<td>0.1042</td>
<td>0.0264</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0520</td>
<td>0.0549</td>
<td>0.0492</td>
<td>0.0827</td>
<td>0.0004</td>
<td>0.8718</td>
<td></td>
</tr>
<tr>
<td>j = 2</td>
<td>0.0774</td>
<td>0.0775</td>
<td>0.0792</td>
<td>0.0740</td>
<td>0.1141</td>
<td>0.0566</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0100</td>
<td>0.0649</td>
<td>0.0003</td>
<td>0.0018</td>
<td>0.0000</td>
<td>0.6170</td>
<td></td>
</tr>
</tbody>
</table>
Table 8
Exchange Rates and Order Flow Driven by Marketwide Private Information in the Primary Smelting and Refining of Nonferrous Metal (except Copper and Aluminum), NAICS 331419

The table shows statistics obtained from the regression

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

In panel A, \( \Delta CUR_{t+j} \) 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, \( H_0: \alpha_l = 0, l > 0 \).

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).

<table>
<thead>
<tr>
<th>Currency</th>
<th>CAD</th>
<th>CHF</th>
<th>FRF</th>
<th>GBP</th>
<th>JPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basket</td>
<td>6.18%</td>
<td>38.21%</td>
<td>2.62%</td>
<td>26.59%</td>
<td>5.10%</td>
</tr>
</tbody>
</table>

**Panel A: Currency Returns**

\( j = 0 \)

\begin{align*}
0.0879 & \quad 0.0872 & \quad 0.0614 & \quad 0.0853 & \quad 0.0435 & \quad 0.0588 \\
0.0212 & \quad 0.4992 & \quad 0.0740 & \quad 0.0383 & \quad 0.3981 & \quad 0.0060 \\
\end{align*}

\( j = 1 \)

\begin{align*}
0.0854 & \quad 0.0933 & \quad 0.0674 & \quad 0.0896 & \quad 0.0429 & \quad 0.0579 \\
0.0156 & \quad 0.1950 & \quad 0.0601 & \quad 0.0173 & \quad 0.4400 & \quad 0.0040 \\
\end{align*}

\( j = 2 \)

\begin{align*}
0.0698 & \quad 0.0858 & \quad 0.0458 & \quad 0.0383 & \quad 0.0417 & \quad 0.0604 \\
0.0761 & \quad 0.1230 & \quad 0.5566 & \quad 0.6835 & \quad 0.2890 & \quad 0.0031 \\
\end{align*}

**Panel B: Excess Currency Returns**

\( j = 0 \)

\begin{align*}
0.0881 & \quad 0.0881 & \quad 0.0631 & \quad 0.0872 & \quad 0.0428 & \quad 0.0565 \\
0.0094 & \quad 0.4571 & \quad 0.0513 & \quad 0.0251 & \quad 0.7840 & \quad 0.0053 \\
\end{align*}

\( j = 1 \)

\begin{align*}
0.0877 & \quad 0.0927 & \quad 0.0710 & \quad 0.0936 & \quad 0.0439 & \quad 0.0574 \\
0.0076 & \quad 0.1593 & \quad 0.0404 & \quad 0.0094 & \quad 0.4311 & \quad 0.0030 \\
\end{align*}

\( j = 2 \)

\begin{align*}
0.0777 & \quad 0.0861 & \quad 0.0471 & \quad 0.0389 & \quad 0.0417 & \quad 0.0600 \\
0.0282 & \quad 0.0975 & \quad 0.5083 & \quad 0.6434 & \quad 0.2920 & \quad 0.0023 \\
\end{align*}
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 in boldface characters.

<table>
<thead>
<tr>
<th>Position ranked</th>
<th>Years 1997</th>
<th>Years 1998</th>
<th>Years 1999</th>
<th>Years 2000</th>
<th>Years 2001</th>
<th>Years 2002</th>
<th>Years 2003</th>
<th>Continuous Top 30 Exporters</th>
<th>2003 Rank</th>
<th>Average Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>331419</td>
<td>331419</td>
<td>331419</td>
<td>331419</td>
<td>331419</td>
<td>315221</td>
<td>315221</td>
<td>315221</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>2</td>
<td>334119</td>
<td>334119</td>
<td>334119</td>
<td>334119</td>
<td>334119</td>
<td>331419</td>
<td>331419</td>
<td>331419</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>3</td>
<td>333999</td>
<td>333132</td>
<td>333999</td>
<td>333132</td>
<td>333999</td>
<td>333999</td>
<td>333999</td>
<td>333999</td>
<td>331419</td>
<td>331419</td>
</tr>
<tr>
<td>4</td>
<td>333132</td>
<td>333999</td>
<td>333132</td>
<td>333132</td>
<td>333132</td>
<td>333132</td>
<td>333132</td>
<td>333132</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>5</td>
<td>333999</td>
<td>333132</td>
<td>333132</td>
<td>333132</td>
<td>333132</td>
<td>333132</td>
<td>333132</td>
<td>333132</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>6</td>
<td>315999</td>
<td>315999</td>
<td>333132</td>
<td>315999</td>
<td>315999</td>
<td>315999</td>
<td>315999</td>
<td>315999</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>7</td>
<td>323119</td>
<td>323119</td>
<td>323119</td>
<td>323119</td>
<td>323119</td>
<td>323119</td>
<td>323119</td>
<td>323119</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>8</td>
<td>325311</td>
<td>325311</td>
<td>325311</td>
<td>325311</td>
<td>325311</td>
<td>325311</td>
<td>325311</td>
<td>325311</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>9</td>
<td>333611</td>
<td>333611</td>
<td>333611</td>
<td>333611</td>
<td>333611</td>
<td>333611</td>
<td>333611</td>
<td>333611</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>10</td>
<td>336413</td>
<td>336413</td>
<td>336413</td>
<td>336413</td>
<td>336413</td>
<td>336413</td>
<td>336413</td>
<td>336413</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>11</td>
<td>335999</td>
<td>335999</td>
<td>335999</td>
<td>335999</td>
<td>335999</td>
<td>335999</td>
<td>335999</td>
<td>335999</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>12</td>
<td>336399</td>
<td>336399</td>
<td>336399</td>
<td>336399</td>
<td>336399</td>
<td>336399</td>
<td>336399</td>
<td>336399</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>13</td>
<td>334513</td>
<td>334513</td>
<td>334513</td>
<td>334513</td>
<td>334513</td>
<td>334513</td>
<td>334513</td>
<td>334513</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>14</td>
<td>322110</td>
<td>322110</td>
<td>322110</td>
<td>322110</td>
<td>322110</td>
<td>322110</td>
<td>322110</td>
<td>322110</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>15</td>
<td>334419</td>
<td>334419</td>
<td>334419</td>
<td>334419</td>
<td>334419</td>
<td>334419</td>
<td>334419</td>
<td>334419</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>16</td>
<td>336111</td>
<td>336111</td>
<td>336111</td>
<td>336111</td>
<td>336111</td>
<td>336111</td>
<td>336111</td>
<td>336111</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>17</td>
<td>334310</td>
<td>334310</td>
<td>334310</td>
<td>334310</td>
<td>334310</td>
<td>334310</td>
<td>334310</td>
<td>334310</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>18</td>
<td>333295</td>
<td>333295</td>
<td>333295</td>
<td>333295</td>
<td>333295</td>
<td>333295</td>
<td>333295</td>
<td>333295</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>19</td>
<td>336120</td>
<td>336120</td>
<td>336120</td>
<td>336120</td>
<td>336120</td>
<td>336120</td>
<td>336120</td>
<td>336120</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>20</td>
<td>315221</td>
<td>315221</td>
<td>315221</td>
<td>315221</td>
<td>315221</td>
<td>315221</td>
<td>315221</td>
<td>315221</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>21</td>
<td>336412</td>
<td>336412</td>
<td>336412</td>
<td>336412</td>
<td>336412</td>
<td>336412</td>
<td>336412</td>
<td>336412</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>22</td>
<td>331211</td>
<td>331211</td>
<td>331211</td>
<td>331211</td>
<td>331211</td>
<td>331211</td>
<td>331211</td>
<td>331211</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>23</td>
<td>334515</td>
<td>334515</td>
<td>334515</td>
<td>334515</td>
<td>334515</td>
<td>334515</td>
<td>334515</td>
<td>334515</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>24</td>
<td>334413</td>
<td>334413</td>
<td>334413</td>
<td>334413</td>
<td>334413</td>
<td>334413</td>
<td>334413</td>
<td>334413</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>25</td>
<td>311212</td>
<td>311212</td>
<td>311212</td>
<td>311212</td>
<td>311212</td>
<td>311212</td>
<td>311212</td>
<td>311212</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>26</td>
<td>323117</td>
<td>323117</td>
<td>323117</td>
<td>323117</td>
<td>323117</td>
<td>323117</td>
<td>323117</td>
<td>323117</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>27</td>
<td>333120</td>
<td>333120</td>
<td>333120</td>
<td>333120</td>
<td>333120</td>
<td>333120</td>
<td>333120</td>
<td>333120</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>28</td>
<td>311223</td>
<td>311223</td>
<td>311223</td>
<td>311223</td>
<td>311223</td>
<td>311223</td>
<td>311223</td>
<td>311223</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>29</td>
<td>325414</td>
<td>325414</td>
<td>325414</td>
<td>325414</td>
<td>325414</td>
<td>325414</td>
<td>325414</td>
<td>325414</td>
<td></td>
<td>331419</td>
</tr>
<tr>
<td>30</td>
<td>333111</td>
<td>333111</td>
<td>333111</td>
<td>333111</td>
<td>333111</td>
<td>333111</td>
<td>333111</td>
<td>333111</td>
<td></td>
<td>331419</td>
</tr>
</tbody>
</table>
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 in boldface characters.

<table>
<thead>
<tr>
<th>NAICS</th>
<th>Number of Firms</th>
<th>Foreign firms</th>
<th>NYSE listed</th>
<th>Incorporated or missing data</th>
<th>Firms that fail to meet the liquidity criterion</th>
<th>Number of available firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>331419</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>333132</td>
<td>16</td>
<td>5</td>
<td></td>
<td></td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>325311</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>336413</td>
<td>11</td>
<td>3</td>
<td></td>
<td></td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>334513</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>336412</td>
<td>6</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>322110</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>334413</td>
<td>27</td>
<td>13</td>
<td></td>
<td></td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>336411</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>