DISCUSSION PAPER SERIES

No. 5604

MARKETWIDE PRIVATE INFORMATION IN STOCKS: FORECASTING CURRENCY RETURNS

Rui Albuquerque, Eva de Francisco
and Luis Marques

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP5604.asp

ISSN 0265-8003

MARKETWIDE PRIVATE INFORMATION IN STOCKS: FORECASTING CURRENCY RETURNS

Rui Albuquerque, Boston University School of Management and CEPR
Eva de Francisco, Congressional Budget Office
Luis Marques, University of Rochester

Discussion Paper No. 5604
April 2006

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5604

April 2006

ABSTRACT

Marketwide Private Information in Stocks: Forecasting Currency Returns*

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. Trades driven by marketwide private information display very little or no correlation with the first principal component of order flow. This finding implies that a simple statistical factor is a poor measure of marketwide private information. Moreover, the model suggests that the previously documented comovement in order flow captures mostly common variation in liquidity trades. We find that marketwide private information obtained from equity market data forecasts industry stock returns and foreign exchange returns consistent with Evans and Lyons' (2004a) model of exchange rate determination.

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

Rui Albuquerque
Boston University School of Management
595 Commonwealth Avenue
Boston, MA 02215
USA
Tel: (1 617) 353 4614
Fax: (1 617) 353 6667
Email: ralbuque@bu.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=147792

Eva de Francisco
Congressional Budget Office
Ford House Office Building
4th Floor
Second and D Streets, SW
Washington, DC 20515-6925
USA
Email: eva.defrancisco@cbo.gov

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=164329

Luis Marques
University of Rochester
Harkness Hall
Rochester, NY 14627
USA
Email: lmrq@troi.cc.rochester.edu

* We thank Marine Carrasco and Werner Ploberger for useful discussions and Clara Vega for many detailed comments on an earlier version of the paper.

Submitted 01 March 2006

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 or official statistics, or from private sources of investors in their use of proprietary models, expertise or insider knowledge. Its content is either asset-specific or aggregate in nature, relating to a group of assets. In this paper, we propose a model of stock trading that allows for the identification of marketwide private information and we investigate the usefulness of marketwide private information across markets.

By definition, aggregate or marketwide private information is useful for trading across a variety of assets.[1] 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.

The paper has two main contributions. The first contribution is to 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 and O'Hara (1992) and 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) aggregate 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 on each information piece. We assume that the arrival of conflicting firm-specific and marketwide private information about the value of a firm is resolved in favor of firm-specific news.

The paper's second main contribution is to show that marketwide private information

[1] We use the terms aggregate private information and marketwide private information interchangeably.

obtained from observed order flow in stocks forecasts equity returns as well as currency returns. The first finding suggests that we are indeed capturing information-driven trading as opposed to non-informative inventory or liquidity effects. In addition, the two findings provide direct evidence consistent with the hypotheses in Evans and Lyons (2004a) that there is marketwide private information and that this private information permeates both equity and currency markets. Evans and Lyons (2004a) focus their analysis on the drivers of the contemporaneous correlation between exchange rate returns and currency order flow in order to address the evidence in Evans and Lyons (2002) and Rime (2001). They predict, but do not test, that order flow in both equity and foreign exchange markets forecast future exchange rate returns.[2]

We implement our model of equity trading on five industries chosen to satisfy two main criteria. First, and with the goal of testing whether marketwide private information can forecast currency returns, these industries have high average ratios of exports relative to total shipments. Our choice is aimed at finding industries for which aggregate 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, but this is not strictly necessary for the analysis; the model partly allows for firm-specific private information to be about the same event as market wide private information, but have different signs across firms in the industry. The second selection criterion is that these industries have multiple firms trading in a liquid fashion in the NYSE.

Our model is estimated with maximum likelihood on microstructure stock trading data from these industries. The cross-section of firms in each industry is used to identify marketwide private information from firm-specific private information and liquidity trades. We use the parameter estimates to construct measures of monthly industry order flow due to aggregate private information. This order flow is the result of artificially decomposing total order flow between order flow driven by aggregate private information from all else in the way dictated by the model.

Estimated marketwide private information has three main properties. First, mar-

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

ketwide private information displays little or no correlation with the total order flow in the industry and with the first principal component in order flow. The later fact implies that marketwide private information cannot be replicated by a simple statistical procedure like a principal component analysis (see Hasbrouck and Seppi (2001)). Instead, our estimate of liquidity trades correlates strongly with the principal component in order flow in the data. Because most of the firms we study are not index constituents, our analysis is not subject to the criticism in Harford and Kaul (undated) that comovement is driven by firms sharing the same index. The second main property of marketwide private information is that it forecasts the stock returns of the firms in the industries we look at. All but one of the 5 industries we study show significant forecasting power up to two months ahead.

The third property of marketwide private information is that it forecasts currency returns. 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 excluding the interest rate differential). Our measures of marketwide private information forecast one and two months ahead currency returns and excess currency returns of the industry-specific baskets with R^2's between 2 and 16 percent. 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 20 percent. We do especially well in forecasting the currencies that are now part of the Euro.

These results complement those in Evans and Lyons (2002, 2004b) in the following way. In these papers there is information content on *currency* order flow to forecast *currency* returns. Their findings are in line with the domestic microstructure literature where a stock's order flow has information content over own *stock* returns, with the difference that Evans and Lyons have in mind marketwide private information, not just firm-specific private information. What we show in this paper is that marketwide private information exists and that it permeates both stock and foreign currency markets. So much so that measures of marketwide private information obtained from order flow in *stocks* forecast *currency* returns.[3]

The role of aggregate private information for microstructure and asset pricing is an open question. The presence of aggregate private information suggests that asset pricing characteristics that may seem anomalous when a market is studied in isolation can be rationalized if one considers simultaneously the portfolio strategies of investors who participate in that and in other markets. Two prominent examples are Gehrig (1993) and Brennan and Cao (1997) who explain the home bias puzzle and the behavior of US investors in international equity markets by implicitly hypothesizing 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. Therefore, demonstrating the existence of aggregate private information validates these and other theoretical studies.[4,5] A potential issue of concern arises if our measures of marketwide private information are truly industry private information that gets diversified away in large portfolios. In such scenario, a rejection of our hypothesis would not necessarily be revealing of the lack of evidence on marketwide private information. On the other hand, finding evidence that marketwide private information forecasts currency returns overthrows this difficulty.

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 present evidence consistent

[3]A complementary analysis is Francis et al. (2005) who study the information-spillover of currency order flow into the equity market finding stronger effects in volatilities than in means.

[4]Examples of other papers where marketwide private information plays a critical role are Subrahmanyam (1991), Chan (1993), Caballe and Krishnan (1994), Kumar and Seppi (1994) and Albuquerque et al. (2004a, 2004b) besides the Evans and Lyons (2004a) paper already cited.

[5]Our study is also related to the empirical literature on foreign currency exposure. This literature explores the relation between exchange rate changes and stock returns of firms in export oriented industries (i.e., with large volume of export revenue to total sales revenue). Evidence for the US and Japanese markets suggests that stock returns of export oriented firms move with exchange rate changes (e.g. Bartov and Bodnar (1994) and He and Ng (1998)). The evidence presented in our paper suggests that these correlations might have their root at least partly in aggregate private information on common economy wide factors such as aggregate productivity.

with the existence of marketwide private information in the Tokyo stock exchange, but do not directly test for it. Albuquerque et al. (2004a) find a common factor in private information from aggregate *net-flow* 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)).

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 currency exposure of the firms in our sample.

2 The Model of Stock Trading

This section presents a model of trading that allows for firm-specific and aggregate private information. The goal of this section is to *identify* the component of observed order flow that is due to private aggregate information from all else using a structural model. It is possible to derive implications for prices and spreads from this model, but because we make no use of this data we omit this analysis from the paper.[6]

2.1 Trading

The model is one of sequential trading where informed and uninformed investors post buy and sell orders to a market maker that sets prices. All agents are risk neutral and competitive. Following Easley and O'Hara (1992) and 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. We deviate from EKOP by allowing investors and the market maker to trade on $I > 1$ risky stocks indexed by $i = 1, ..., I$, 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 θ, there can be an information event containing

[6]The details are available from the authors upon request.

aggregate private information which is useful across all I assets. Such event contains good (bad) private aggregate information news with probability $1-\rho$ (ρ) and affects positively (negatively) all I assets. Second, regarding firm i, with probability α_i, there can be 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 $1-\delta_i$ (δ_i). The arrival of aggregate 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.

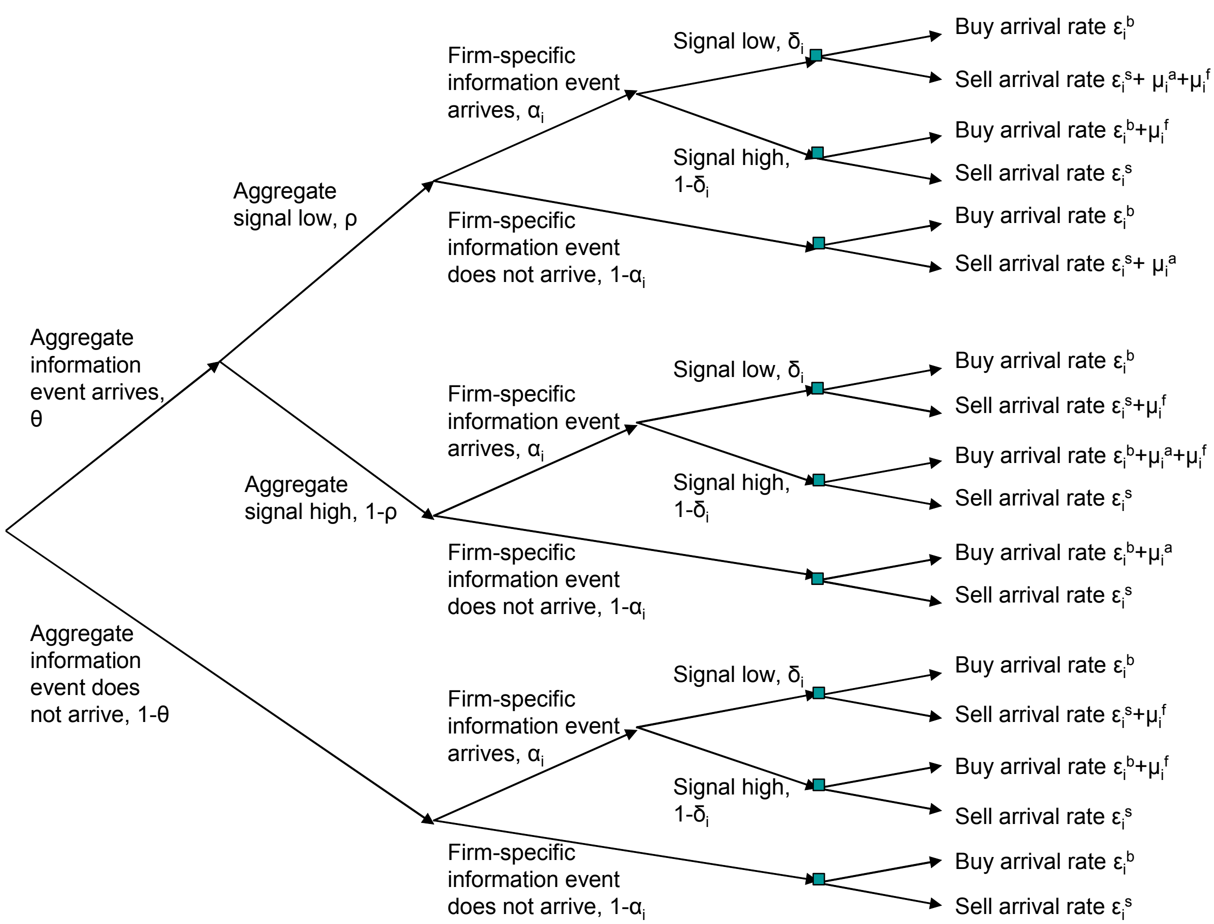

Figure 1: Tree diagram of the trading process for stock i.

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 aggregate private information and firm-specific private information, only the former, or only the later. With more than one piece of news affecting trading we adopt a simple rule for conflict resolution: whenever aggregate and firm-specific private information on any firm i are qualitatively contradictory, the firm-specific news dominates investors' behavior. Informed investors may therefore not act in accordance to their aggregate private information.

Consider for example a day with good aggregate 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, let μ_i^f be the average arrival rate of informed investors who trade based on firm-specific private information news and μ_i^a be the average arrival rate of informed investors who trade based on marketwide private information news.[7] If on any day the qualitative nature of firm-specific and marketwide private information coincides, then the average arrival rate of informed investors is $\mu_i^f + \mu_i^a$. Together with informed investors, uninformed investors buy orders arrive at an average rate ε_i^b and sell orders arrive at an average rate ε_i^s. These parameters are constant 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 aggregate and firm-specific private information news at the top of Figure 1. The total volume of sell orders has an average of $\mu_i^f + \mu_i^a + \varepsilon_i^s$ whereas only uninformed investors buy so the total volume of buy orders has an average of ε_i^b. At the bottom of the tree the event with no private aggregate information and no firm-specific

[7] We let μ_i^a vary by firm as each firm might have different sensitivities to aggregate factors.

private information on firm i generates trading only by uninformed investors. Thus, the average of buy orders is ε_i^b and that of sell orders is ε_i^s. The rest of the arrival rates at the end of each node is constructed in similar fashion.

2.2 The Likelihood Function

We now construct the likelihood function of observing $\left\{(S_{it}, B_{it})_{t=1}^T\right\}_{i=1}^I$ sell and buy orders respectively, for I firms during T trading days. A day with good aggregate private information news occurs with probability $\theta(1-\rho)$. Thus, given good aggregate private information, the conditional probability of observing the pair of sell and buy orders $\{S_i, B_i\}$ on firm i is

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

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

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

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

The probability of observing the buy and sell orders $\{S_{it}, B_{it}\}_{i=1,\ldots,I}$ on day t of bad aggregate private news is $\Pi_{i=1}^I l_B(\{S_{it},B_{it}\})$.

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

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

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

We can now construct the likelihood function of the data. On any day t the unconditional likelihood of observing I buy orders $\{B_{it}\}$ and I sell orders $\{S_{it}\}$ is the weighted average of the expressions above, with the weights given by the probability of each type of aggregate information event

$$\begin{aligned} l\left(\left(S_{it}, B_{it}\right)_{i=1,\ldots,I}\right) &= \theta\left(1-\rho\right) \Pi_{i=1}^I l_G\left(\{S_{it}, B_{it}\}\right) \\ &+\theta\rho \Pi_{i=1}^I l_B\left(\{S_{it}, B_{it}\}\right)+\left(1-\theta\right) \Pi_{i=1}^I l_0\left(\{S_{it}, B_{it}\}\right). \end{aligned}$$

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

$$L\left(\left(S_{it}, B_{it}\right)_{t=1,\ldots,T; i=1,\ldots,N}\right)=\Pi_{t=1}^T l\left(\left(S_{it}, B_{it}\right)_{i=1,\ldots,I}\right). \tag{1}$$

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

Allowing the parameters μ^f, μ^a, and ε^b and ε^s to vary with i gives the model flexibility to capture different trading intensities to news across firms. To estimate the parameters that drive the release of firm-specific private information (α_i, δ_i), we need 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 aggregate private information the 'daily' likelihoods are trinomials of Poisson probability functions which are bilinear in α_i and δ_i (see Easley et al. (1997) and Vega (2004)). Based on a model of trading of individual stocks in isolation, Vega (2004) argues that $\left(\varepsilon_i^s, \varepsilon_i^b\right)$ measure the average number of sell orders and the average number of buy orders for firm i, and that μ_i^f measures the abnormal number of buy or sell orders that are firm-specific. With some qualifications addressed next, this is true too in our model.

In our model, μ_i^a also captures abnormal trading, but only that abnormal trading which is observed as a pattern across several firms in a single day. The richness of possible events in the model permits estimation of μ_i^a separately from μ_i^f: when private aggregate 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 (θ, ρ).

In contrast to Easley et al. (1997) and Vega (2004), 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 α_i, δ_i, μ_i^a, μ_i^f, ε_i^s and ε_i^b depends on the estimation of the other firms' parameters as they are linked by the arrival of aggregate private information news. To see this consider solving a maximum likelihood problem as in Easley et al. (1997) and Vega (2004) –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 α_i, ε_i^s and ε_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 μ_i^a is sufficiently small, leading to an upward bias in $\hat{\varepsilon}_i^s$ and $\hat{\varepsilon}_i^b$.[8] If, in contrast, the true μ_i^a is close to the true μ_i^f, a day with aggregate private information will be mistaken for a day with firm-specific private information, biasing the value of $\hat{\mu}_i^f$ downwards (upwards) if $\mu_i^f > \mu_i^a$ ($<$) and affecting the inference of α_i and δ_i as well. It is easy to construct more 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

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

solution).

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

$$\text{Average Informed (Agg.) Buys} = \theta\rho\sum_{i=1}^{I}\left[1-\alpha_i+\alpha_i(1-\delta_i)\right]\mu_i^a. \tag{2}$$

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

$$\text{Average Informed (Agg.) Sells} = \theta(1-\rho)\sum_{i=1}^{I}(1-\alpha_i+\alpha_i\delta_i)\mu_i^a. \tag{3}$$

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

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

We make use of the variable MPI to capture the qualitative nature of the aggregate 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 aggregate 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 and instead lumps them together. This is fine if one wishes to forecast absolute returns, or if the focus is on the quantity of information asymmetry, for example, to measure information risk in a stock, 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 aggregate 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 changes. In their model, the presence of transitory and persistent productivity shocks, means that both flows into the equity market and the currency market are correlated with and forecast changes in the exchange rate.[9] This follows because in their model investors trade based on their private information, but set prices only based on public information. After trading, each investors' private information becomes public, as order flow is observable, and exchange rate quotes are revised subsequently.

[9] Evans and Lyons (2004b) document that *foreign exchange* order flow has predicitve 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. As we view it ours is a more direct test of the model.

In Evans and Lyons (2004a), because order flow is observable it can only forecast one-period ahead currency returns (afterwards the information contained in order flow is immediately incorporated into prices and everyone's decisions.) However, in real economies order flow is not observable and private information is not common across agents in each country. 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. 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.

Hypothesis 2 above makes a qualification on the predictability of currency returns; we are not interested in all private information driven trades, but only those trades that refer to aggregate 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 aggregate private information from trades due to firm-specific private information and liquidity trades.

3 Data

In this section we first describe our industry and firm selection, 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 international exposure. Specifically, we focus on industries with significant export sales relative to total shipments.

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). Because our data spans the period from January 1993 to December 2003, we restrict attention to industries for which there is a complete bridge between the 2002 version of NAICS and its predecessor, the Standard Industrial Classification code or SIC.[10] We then selected those industries with a significant international exposure. 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.[11] The industries we pick rank consistently among the top 20 in exports to total shipments over the 1993-2003 period. We also make use of this 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 further restrict the set of possible industries to those industries with at least two firms, where only firms traded in the NYSE were considered. 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 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 private aggregate information θ and the fraction of time aggregate news are bad news ρ; and (*ii*) because the number of degrees of freedom increases with the number of firms (see below).

We now turn to the selection of firms included in the analysis. First, only companies with data available from both the Trade and Quotes (TAQ) database – for order flow data

[10] In 1997, the Office of Management and Budget adopted NAICS, a system for classifying establishments by type of economic activity, to replace the 1987 Standard Industrial Classification (SIC) codes. 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.

[11] Available at http://dataweb.usitc.gov/scripts/user set.asp and http://www.census.gov/mcd/asmas2.html, respectively.

– and CRSP/Compustat database – for industry classification information – are included. Second, we impose a month-by-month liquidity constraint on each firm which guarantees a minimum of 7 trades on average per day in any given month (EKOP, 1996). Third, with the purpose of identifying firms with qualitatively similar exposures to exchange rates, we excluded foreign firms that did not have a significant presence in terms of operations in the US. For example, this criterion lead us to include the U.K. firm Doncasters PLC in the Aircraft Engines Manufacturing sector and to exclude the Brazilian firm Embraer from the Aircraft Manufacturing sector.[12]

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 aggregate 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 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

[12]Even within industries with significant foreign exchange exposure, firms can differ in their qualitative exposure to the same exchange rates or can be exposed to different exchange rates. The model of trading in section 2 allows for this possibility somewhat: firm-specific private information can be contemporaneous with marketwide private information, but have different signs for different firms in the industry. However, the model does not allow for a correlation between the arrival of firm-specific and marketwide private information as their arrival rate is i.i.d.

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 proposed a 5-second rule for comparing execution prices to quotes reported a minimum of 5 seconds before the transaction was reported.[13]

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 mostly from Datastream and is complemented with data from the International Financial Statistics of the International Monetary Fund. We use *beginning of month quotes* of foreign currency (FC) per US dollar (USD), denoted as $S_{FC/\$,t}$. Currency returns in month t, cr_t, are given by

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

With our notation, a positive cr represents an appreciation of the USD relative to the FC. 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 returns using as weights the previous month's fraction of industry exports going to each

[13] 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.

country. In any one 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 one (see below). Monthly data for exports is free alongside ship value of domestic exports and for imports is customs value imports for consumption from the US Trade Commission database.

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. 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)). This does not mean that U.S. 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. We leave to the appendix a brief description by sector of currency exposure reported in financial statements.

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.[14] The companies (and dates) included in the analysis are Baker Hughes (1/93-12/03), Weatherford International (9/98-12/03), Varco International (1/93-12/03), CAMCO International (12/93-8/98), Cooper Cameron Corp. (8/95-12/03), National-Oilwell (10/96-12/03), IRI International (11/97-6/00), NATCO Group (1/00-12/03), Grant Prideco (4/00-12/03), Oil States International (2/01-12/03), and FMC Technologies (6/01-12/03). This is the sector for which we have most firms and therefore for which we expect to obtain the best results.

[14] Definitions available from `http://www.census.gov/epcd/naics02/naico602.htm#N31`.

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 events.[15] On average total imports represented 96 percent of total shipments between 1997 and 2001. From the US International Trade Commission database we learn the currencies of the countries most relevant in terms of import shares: Canadian dollar, British pound, and all major currencies recently replaced by the euro. To compute the import-weighted exchange rate we make use of the following currencies depending on their relevance:[16] 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.[17]

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 (1/93-9/03), Gulfstream Aerospace (1/96-7/99), Grumman Corp. (1/93-4/94) and McDonnell Douglas Corp. (1/93-7/97).[18] We exclude from the analysis the period 8/99-9/03 where only Boeing is present.

The dominant currencies associated with this industry are the British pound, Japanese yen, and several European currencies. On aggregate, from 1997 to 2002 exports were on average 48 percent 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

[15] However, this sector's MPI does a good job at predicting the Mexican peso and the Indonesian rupiah, but not the Russian ruble.

[16] Recall that at any month there are at most 5 currencies in the basket.

[17] 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.

[18] Whitehall Corporation is part of this sector but never meets the liquidity criterion. Embraer is another company from this sector listed on the NYSE but it was excluded from the sample on the grounds of not being American and not having a significant presence in the United States in terms of operations.

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 (1/99-12/03), Sequa Corp (1/93-12/03), UNC Inc (1/93-7/97), United Technologies Corp (1/93-12/03), Doncasters Plc (1/97-7/01), and Howmet International Inc (11/97-6/00), all of which are incorporated in 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 percent of all exports in our sample. On average exports from this sector were about 50 percent 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 (1/93-12/03), Sundstrand Corp. (1/93-5/99), Talley Industries (1/93-1/98), Triumph Group (10/96-12/03), Rockwell Collins Inc. (7/01-12/03), Goodrich Corp. (1/93-12/03), Rohr Industries (1/93-12/97), and Ducommun Corp. (11/96-12/03).

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 are British pound, Japanese yen, and the Canadian dollar. Between 1997 and 2001, exports as a fraction of total shipments averaged 57 percent for the sector.

Primary Smelting and Refining of Nonferrous Metal (except Copper and Aluminum): NAICS 331419. This sector 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. (1/93-1/03), WHX (1/93-12/03), INCO (1/93-12/03), and WMC/Alumina Corp. (1/93-12/03). Of these, INCO is incorporated in Canada and WMC/ Alumina is incorporated in Australia. We include these two foreign incorporated firms because both firms have significant operations in the US. Moreover, INCO, in spite of being Canadian, uses the USD as its functional and reporting currency. However, as we will see below, the measure of MPI that is obtained in this sector is quite different from all others with implications for the results on forecasting currency returns, but mainly equity returns. Alternatively, we have repeated the analysis just with the US firms and just with the foreign firms (not reported). These alternatives do better in forecasting currency returns, but fail still in forecasting equity returns, we believe because of the few degrees of freedom in estimating MPI when only two firms are used.

The main outputs in this industry are smelted nickel (used as input in the stainless steel industry) and titanium. 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 percent of all exports from the sector. All other currencies have much smaller weights in terms of trade and are usually not included more than once in the basket. Average total exports over the period 1997-2001 represented approximately 190 percent of total shipments.

4 Results

In this section we start by giving details on the maximum likelihood estimation in (1) and the estimation of MPI. Then we lay out the regression specifications, exact hypotheses tests, and 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 sector, we estimate the value of marketwide private information in every month t MPI_t, by using the vector of estimated parameters $\hat{\Theta}_t \equiv \left(\hat{\alpha}_i, \hat{\delta}_i, \hat{\theta}, \hat{\rho}, \hat{\varepsilon}_i^s, \hat{\varepsilon}_i^b, \hat{\mu}_i^a, \hat{\mu}_i^f\right)_{i=1,\ldots,I}$ that maximizes the likelihood function (1) jointly across all firms in month t. This is a constrained maximization as the probabilities $0 \leq \alpha_i, \delta_i, \theta, \rho \leq 1$ and the trading intensities $0 \leq \hat{\varepsilon}_i^s, \hat{\varepsilon}_i^b, \hat{\mu}_i^a, \hat{\mu}_i^f$. With N firms and 22 trading days per month, we have a total of $N \times 22$ observations to estimate $N \times 6 + 2$ parameters, yielding $N \times 16 - 2$ degrees of freedom. The quality of the estimation therefore increases as the number of firms increases. 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 $\left(\hat{\alpha}_i, \hat{\delta}_i, \hat{\varepsilon}_i^s, \hat{\varepsilon}_i^b, \hat{\mu}_i^a, \hat{\mu}_i^f\right)$. 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 aggregate private information news (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 μ_i^f and μ_i^a. To minimize the impact of initial conditions we proceed by first implementing, *for every sector and month*, a grid search on the parameters μ_i^f and μ_i^a. 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 ε_i^s and ε_i^b to each month's mean daily value of sell and buy orders, respectively (Vega (2004)). We set the initial values of the remaining parameters θ, ρ, α_i and δ_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.[19] This procedure is applied over all months and sectors for consistency. While this results in a very time consuming procedure, especially because gradient search methods were of no use here, it makes us more confident that our parameter estimates attain the global maximum.

Performing this estimation for each month and industry means we need not worry about possible non-stationarity of information with respect to the parameters μ_i^f, μ_i^a, ε_i^s, and ε_i^b. However, the observed growth in trading activity over the years requires that we detrend our measures of MPI. Easley et al. (2002) account for this non-stationarity endogenously in their estimation. With many firms this is cumbersome and costly to implement. Hence, we take the shortcut of using the Hodrick-Prescott filter on our estimate of order flow driven by private aggregate information. The growth in trading activity and the fact that our firms are generally large firms with a large volume of trading also means that estimating MPI for the final months of available data was increasingly difficult. Specifically, for some months the number of buys and/or sells was so high that maximizing the likelihood function required values higher than the largest positive floating-point number in our personal computers. For this reason the sample length varies across sectors (see Table 1 below).

4.2 MPI and Factors in Equity Order Flow

Table 1 describes some of the properties of MPI by correlating it to other data and model variables. We use this information to provide a characterization of MPI.

Table 1 shows that MPI is positively correlated, but only weakly so for 3 of the sectors, with total industry order flow, TOF (i.e., the sum of total buy orders minus the sum of total sell orders over the month). For two industries this correlation is statistically significant at the 10 percent significance or better. In addition, because Hasbrouck and Seppi (2001) have shown that there is significant comovement in order flows of 30 stocks in the Dow Jones Industrial, it is natural to ask whether a simple factor analysis gives a measure of marketwide private information that can be used in place of the more

[19] The implemention of the EM algorithm, that is more robust to initial conditions, did not improve results and slowed down estimation considerably.

complex measure we derive from a structural model. To answer this question we extract a principal component from firm-level order flow in each month and industry, $PC1$. A large first principal component indicates significant comovement in equity order flow. The first principal component (linear combination of firms' order flow) accounts for a significant portion of the total intra-month variability of order flow across industries.[20] Numbers range from 61.5 percent in industry 333132 to 95.4 percent in industry 336412 (untabulated), and in contrast we find that the percentage of explained variation in total order flows from MPI is below 10 percent for all industries.[21] Not surprisingly, Table 1 shows that the correlation between $PC1$ and TOF is almost equal to 1.

Table 1 indicates that the correlation between MPI and the first principal component $PC1$ (both detrended) is low and statistically insignificant in most industries. This fact turns out not be too surprising because MPI excludes any common variation in order flow across firms in the same industry that is due to common variation in liquidity trades (ε_i^b and sells ε_i^s). Table 1 also reports the correlation of estimated aggregate liquidity trades $LT_t = \sum_{i=1}^{N} \left(\hat{\varepsilon}_{it}^b - \hat{\varepsilon}_{it}^s\right)$ and MPI and $PC1$. While LT is strongly positively associated with $PC1$ it displays a negative correlation with MPI. We conclude that the principal component is capturing some of the common variation due to aggregate private information but mostly any common variation due to residual liquidity trades (possibly from portfolio-wide shocks). Documenting the existence of marketwide private information as we do here via a structural model takes the statistical exercise in Hasbrouck and Seppi (2001) one step further arguing that perhaps a strong component of what is captured with a principal component of order flow is due to liquidity trades (see subsection 4.5 below for more on MPI and $PC1$). The finding that LT has a very strong and negative correlation with MPI is unexpected as in the theory the arrival rates of uninformed trades are independent of the arrival of marketwide private information. It appears that the data contains even more correlation in order flow across firms than what MPI is able to capture. Because we show below that MPI is able to forecast stock returns, we

[20] To perform the principal components analysis we used Matlab's function `princomp`. This function returns the principal component coefficients which are then used as weights for each firms' order flow.

[21] The percentage of explained variation in total order flow accounted for by MPI can be computed from line 1 in Table 1 by squaring the correlation numbers.

are confident that what we estimate is information driven, but recognize that it may be a lower bound of the true marketwide private information trading. The study of such additional correlation is an interesting question for future research.

The sum of MPI and LT is highly correlated with total order flow TOF, but it is not a model average of total trading activity. For that we need the model's estimate of trading due to firm-specific private information, $FPI_t = \sum_{i=1}^{N} \hat{\alpha}_{i,t} \left(1 - 2\hat{\delta}_{i,t}\right) \hat{\mu}_{i,t}^{f}$. As with MPI, FPI is highly negatively correlated with liquidity trading, but remarkably the sum $MPI + FPI + LT$ is highly correlated with TOF (correlations around 0.97 for three of the sectors, and 0.74 and 0.37 for the other two) and with $PC1$. The high correlation between $MPI + FPI + LT$ and TOF gives us confidence that, at least for the first 4 sectors, the model appears to be doing a good job decomposing order flow. For sector 331419, the correlation is 0.37 less than half of that in the other sectors.

There is a final aspect of Table 1 that we would like to highlight. The last column gives the numbers for sector 331419. It is apparent from looking across sectors that this sector's MPI stands out and contrasts with the properties of MPI for the other sectors. We believe that this is related to the fact that this sector uses four firms to estimate MPI, two American and two foreign firms (see subsection 3.3). While there are no good fixes – dropping two of the firms significantly reduces the degrees of freedom – we proceed with the analysis under this caveat.

As a final comment on MPI and comovement we note that Harford and Kaul (undated) 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 331132 are index constituents, only 1 out of 6 firms in NAICS 336411 belonged to an index, only 1 out of 6 in NAICS 336412 belonged to an index, only 4 out of 8 firms in NAICS 336413 belonged to an index, and of the 4 companies in NAICS 331419 none was an index constituent.[22]

4.3 Specification of the Forecasting Regressions and Hypotheses Tests

4.3.1 Forecasting Equity Returns

For each sector, we regress firm stock returns on lagged values of MPI as indicated by the regression model:

$$RET_{i,t+j} = a_{i,0} + \sum_{l=1,\dots,L} a_l MPI_{t-l} + u_{i,t+j}, \tag{4}$$

where in each regression $RET_{i,t+j}$ is either the j-month ahead holding period stock return 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.[23] 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. The regressions are conducted with contemporaneous ($j = 0$), one month ahead ($j = 1$) and two months ahead ($j = 2$) equity returns. Hypothesis 1 is that MPI forecasts equity returns:

$$H_0 : a_1 = ... = a_L = 0,$$

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 (4), 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

[22] 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.

[23] 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 signficance tests.

estimator.[24] 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 that we are certain 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 sector, we report results for two regression specifications using either currency returns or excess currency returns. Let ΔCUR_{t+j} be either the value of the currency return at time $t+j$ or the value of the excess currency return at time $t+j$. Our regression model is:

$$\Delta CUR_{t+j} = \alpha_0 + \sum_{l=1,\ldots,10} \alpha_l MPI_{t-l+1} + u_{t+j}. \tag{5}$$

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 (Hypothesis 2) is that MPI can explain currency movements:

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

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

In specification (5), 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 (5) where $\alpha_2 = ... = \alpha_{10} = 0$. This specification presumes that private information diffuses quickly and is fully incorporated

[24] STATA's generalized Hausman test procedure rejected the random effects specification in every case as well as the random coefficients specification.

into prices in one month. Instead, the specification in (5) assumes that new information about future fundamentals summarized by the component of order flow associated with aggregate 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 (5) 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)).[25] 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 $\alpha_l = 0$.[26] Since we are interested in testing $H_0 : \alpha_l = 0$, $l = 1, ..., L$, we have that, under the null, the uncorrected Newey-West estimator of the asymptotic variance of the OLS estimator of regression coefficients in (5) remains consistent.

4.4 Forecasting Results

We start with the results for equity returns in Table 2. Each cell of Table 2 has three numbers: the first is the sum of the estimated coefficients $\sum_{l=1}^{L} \hat{a}_l$, below it is the $p-$value on the significance of the sum of the coefficients and the third is the $p-$value on the null hypothesis $H_0 : a_l = 0, l = 1, ..., L$. Our estimated information shock, MPI, successfully forecasts returns in 4 out of the 5 sectors. For sectors 333132, 336411, 336412 and 336413, not only do we reject the null of no explanatory power of MPI, but we also

[25] An alternative is to use bootstrapping. In this case, however, this is not feasible as it would require bootstrapping not only the second step estimation, but also *simultaneously* the first very-time-consuming MLE step.

[26] We use robust standard errors because of the autocorrelation induced by overlapping observations (see Hansen and Hoddrick (1980)).

get positive significant estimates for the cumulative response of equity returns.[27] Only for the one month ahead stock return regression of NAICS 336412 we get a globally significant regression, but with a sum of coefficients associated with current and lagged MPI not statistically significant. On the negative side, we cannot forecast returns in sector 331419 with MPI. While we are not certain why this failure occurs, we note our discussion above on how this sector may have trouble in generating a reasonable estimate of marketwide private information (subsection 4.2). We also note that this sector 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 sectors) and the highest kurtosis (11.25 versus 4.8-7 for the other sectors) making it harder for MPI to explain returns.

Now turn to the forecasting of currency returns in Tables 3-7.[28] The summary of our results is simple: MPI forecast currency returns as well as excess currency returns quite well: (i) MPI appears to forecast simple currency-basket returns better than excess 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; and (iv) the R^2's in these regressions are in the range of $5 - 25$ percent.

Panels A and B of Table 3 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

[27] 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 volume of trade 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)).

[28] Recall that in these tables the individual currencies that we forecast 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, if outside these devaluation periods.

described in subsection 4.3 above. (This structure is repeated in Tables 4-7.) Because of the large number of firms with apparent similar exposures in this sector, we expect that this is one of the sectors where MPI is better estimated. Indeed, we find that the R^2's in the forecasting regressions in this sector 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 sector.

Table 4 gives the results for the Aircraft Manufacturing sector (NAICS 336411). MPI shows significant predictive power for most currencies and specifications. In fact, of all the currencies selected, only for contemporaneous and one month ahead returns of the British pound and for raw returns of the currency basket do we not find any significant correlation with MPI. In particular, for the Canadian dollar and the Japanese yen we find R^2's in excess of 20 percent and in most instances the regressions are globally significant at the 1 percent significance level.

Table 5 displays the results for the Aircraft Engine Manufacturing sector (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 sector, 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.[29]

Table 6 shows the results for the Aircraft Parts Manufacturing sector (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 sectors, MPI accounts for a somewhat smaller share of each currency return's total variation: overall R^2's are somewhat lower.

[29] Interestingly, contemporaneous MPI alone can forecast one month ahead returns of the German mark. The fact that a distributed lag structure of MPI cannot, is probably because of either the overfitting of the regression – which creates noise and a loss of power – or of the loss of observations due to the inclusion of the 9 lags. But this is clearly the exception.

Finally, Table 7 presents the results for the Primary Smelting of Nonferrous Metal industry (NAICS 331419). Contemporaneous and lagged MPI forecast the currency basket in terms of raw and excess returns at almost every horizon. This is in spite of the difficulty in forecasting stock returns in this sector. 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 first sectors.

The lower ability to forecast currency returns in sector 331419, in particular of the Canadian dollar and of the British pound, derives from a noisier estimate of marketwide private information from having a sector composed of 4 companies, two American, one Canadian and one Australian. To assess this hypothesis we have redone the analysis only for the two foreign firms. Despite the obvious lack of degrees of freedom, untabulated results confirm this suspicion. MPI extracted from the order flow of the two foreign firms alone correlates contemporaneously with the Australian dollar and the Canadian dollar and forecasts the one month ahead returns on the Swiss franc.

4.5 Robustness Checks

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

Table 1 documented 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 percent 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 that this statistical factor captures all common variation in order flow across firms, including presumably that 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 (provided upon request) show that MPI does not lose explanatory power with the introduction of $PC1$. In every sector, contemporaneous and lagged MPI remain jointly significant (whenever this was the case before) whereas $PC1$ only does so for sectors 336413 and 331419, and to a lesser extent for sector 333132. (Recall that sectors 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 percent (less that 1 p.p.) on average for sector 331419 and about 8 percent (or 1.2 p.p.) for sector 333132. For the remaining sectors 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 can only conclude that $PC1$ is capturing some other source of explanatory power. 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.[30] The results (provided 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 sector 336412 and for the raw and excess returns on the British

[30] 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.

pound do we find that, at the optimal AR lag length, the current and lagged MPI ceases to be significant.

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 aggregate private information is not well captured by a simple statistical factor of order flow (Hasbrouck and Seppi (2001)). Moreover, using our structural model, estimated liquidity trades display a strong positive association with the first principal component in order flow.

We conduct our analysis with the intent of showing that marketwide private information permeates both the equity and currency markets. Accordingly, 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 model of exchange rate determination in Evans and Lyons (2004a). The finding of marketwide private information is also critical to many models of asset pricing that assume the existence of private information at an index level, i.e. aggregate private information.

We view as an important step for future research the study of alternative approaches that simplify the computational burden involved in structurally estimating marketwide private information.

Appendix

This appendix provides some more details on currency exposure that is explicitly mentioned by the firms in our study in their annual reports and 10-K forms.

Oil and Gas Field Machinery and Equipment Manufacturing: NAICS 333132. In 2003 Baker Hughes reports having entered foreign currency forwards to 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. Cooper Cameron has production facilities located in the United Kingdom and other European and Asian countries. To the extent it invoices in foreign currency, 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.

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 percent of total sales, 40 percent 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 is not nill; its competitiveness is affected by USD movements.

Aircraft Engine and Engine Parts Manufacturing: NAICS 336412. Sequa reports primary foreign currency exposure to the British pound and the euro. Howmet

International had during the sample period in analysis operations in France, the United Kingdom, Canada, and Japan. It also had forward exchange rate contracts in British pound, French franc, and Japanese yen. United Technologies had at a point over 50 percent of its revenue coming from foreign markets, namely Europe and Asia-Pacific.

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, and the Canadian dollar. Sundstrand Corp. acknowledges exposure to fluctuations in foreign currencies for transactions denominated primarily in the British pound, French franc, and Singapore dollar.

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, INCO says in its 2003 annual report that notwithstanding the use of foreign currency forwards on the Canadian dollar, the Euro and the Australian dollar, changes in exchange rates can have a 'material' impact on future earnings and cash flows. 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 percent of its sales revenue originated in Europe, of which 60 percent was denominated in currencies other than the USD, mainly the British pound and European currencies now in the Euro area.

References

[1] Albuquerque, Rui, Greg Bauer, and Martin Schneider (2004a), "Global Private Information in International Equity Markets," working paper University of Rochester.

[2] Albuquerque, Rui, Gregory H. Bauer, and Martin Schneider (2004b), "International Equity Flows and Returns: A Quantitative Equilibrium Approach," working paper University of Rochester.

[3] Barclay, Michael J., Robert H. Litzenberger, and Jerold B. Warner (1990), "Private Information, Trading Volume, and Stock-Return Variances," *Review of Financial Studies* 3: 233-253.

[4] Bartov, Eli, and Gordon M. Bodnar (1994), "Firm Valuation, Earnings Expectations, and the Exchange-Rate Exposure Effect," *Journal of Finance* 49: 1755-1785.

[5] Bauer, G. H., and C. Vega (2004), "The Monetary Origins of Asymmetric Information in International Equity Markets," working paper University of Rochester.

[6] Brennan, Michael J., and H. Henry Cao (1997). "International Portfolio Investment Flows," *Journal of Finance* 52, 1851-1880

[7] Caballe, J., and M. Krishnan (1994), "Imperfect Competition in a multi-security market with risk neutrality," *Econometrica* 62: 695-704.

[8] Chan, Kalok (1993). "Imperfect Information and Cross-Autocorrelation among Stock Prices," *Journal of Finance* 48, 1211-1230.

[9] Easley, David, and Maureen O'Hara (1992), "Time and the Process of Security Price Adjustment," *Journal of Finance* 47: 577-605.

[10] Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman (1996), "Liquidity, Information, and Infrequently Traded Stocks," *The Journal of Finance* 51: 1405-1436.

[11] Easley, David, Nicholas M. Kiefer and Maureen O'Hara (1997), "One Day in the Life of a Very Common Stock," *The Review of Financial Studies* 10: 805-835.

[12] Easley, David, Robert F. Engle, Maureen O'Hara and Liuren Wu (2002), "Time-Varying Arrival Rates of Informed and Uninformed Trades," working paper Cornell University.

[13] Ellis, Katrina, Roni Michaely, and Maureen O'Hara (2000), "The Accuracy of Trade Classification Rules: Evidence from Nasdaq," *Journal of Financial and Quantitative Analysis* 35: 529-551.

[14] Evans, Martin D. D., and Richard K. Lyons (2002), "Order Flow and Exchange Rate Dynamics," *Journal of Political Economy* 170-180.

[15] Evans, Martin D. D., and Richard K. Lyons (2004a), "A New Micro Model of Exchange Rate Dynamics," working paper UC Berkeley.

[16] Evans, Martin D. D., and Richard K. Lyons (2004b), "Exchange Rate Fundamentals and Order Flow," working paper UC Berkeley.

[17] Francis, Bill B., Iftekhar Hasan, and Delroy M. Hunter (2005), "Dynamic Relations between International Equity and Currency Markets: The Role of Currency Order Flow," *Journal of Business*, forthcoming.

[18] Gehrig, Thomas (1993), "An Information Based Explanation of the Domestic Bias in International Equity Investment," *Scandinavian Journal of Economics* 95: 97-109.

[19] Goldberg, Linda S., and Cédric Tille (2004), "Vehicle Currency Use in International Trade," working paper Federal Reserve Bank of New York.

[20] Hansen, Lars, and R. J. Hodrick (1980). "Forward Exchange Rates as Optimal Predictors of Future Spot Exchange Rates: an Econometric Analysis," *Journal of Political Economy*, 88, 829-853.

[21] Harford, Jarrad, and Aditya Kaul (undated), "Correlated order flow: Pervasiveness, sources and pricing effects," *Journal of Financial and Quantitative Analysis* forthcoming.

[22] Hasbrouck, Joel (1991), "Measuring the Information Content of Stock Trades," *Journal of Finance*, 46, pp. 179-207.

[23] Hasbrouck, Joel, and Duane Seppi (2001), "Common factors in prices, order flows, and liquidity," *Journal of Financial Economics* 59: 383-411.

[24] He, J., and L.K. Ng (1998), "The foreign exchange exposure of Japanese multinational corporations," *Journal of Finance* 53: 733-753.

[25] Kumar, Praveen, and Duane J. Seppi (1994), "Information and Index Arbitrage," *The Journal of Business* 67: 481-509.

[26] Lee, Charles M.C., and Mark J. Ready (1991), "Inferring Trade Direction from Intraday Data," *Journal of Finance* 46: 733-746.

[27] Lee, Charles M. C., and Balkrishna Radhakrishna (2000), "Inferring investor behavior: Evidence from TORQ data," *Journal of Financial Markets* 3: 83-111.

[28] Lewellen, Jonathan (2004). "Predicting Returns with Financial Ratios," *Journal of Financial Economics* 74, 209-235.

[29] Newey, Whitney K., and Daniel McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in Z. Griliches and M.D. Intriligator (Eds.), *Handbook of Econometrics*, Vol. 4, Chapter 36, pp. 2111-2245.

[30] Odders-White, Elizabeth R. (2000), "On the Occurrence and Consequences of Inaccurate Trade Classification," *Journal of Financial Markets* 3: 259-286.

[31] Pagan, Adrian (1984), "Econometric Issues inthe Analysis of Regressions with Generated Regressors," *International Economic Review* 25: 221-246.

[32] Rime, Dagfinn (2001), "Private or Public Information in Foreign Exchange Markets? An Empirical Analysis," working paper Central Bank of Norway.

[33] Stambaugh, Robert F. (1999). "Predictive Regressions," *Journal of Financial Economics* 54, 375-421.

[34] Subrahmanyam, Avanidhar (1991), "A Theory of Trading in Stock Index Futures," *Review of Financial Studies* 4, 17-51.

[35] Vega, Clara (2004), "Stock Price Reaction to Public and Private Information," forthcoming *Journal of Financial Economics.*

[36] Yu, Lei (2005), "Basket Securities, Price Formation, and Informational Efficiency," working paper University of Notre Dame.

Table 1
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.

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

Table 2
Stock Returns and Order Flow Driven by Market-wide Private Information.

The table shows statistics obtained from the regression

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

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 aggregate or 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 $\sum_{l=1,\ldots,L} a_l$, 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.

Coefficients and p-values	j	NAICS 333132	NAICS 336411	NAICS 336412	NAICS 336413	NAICS 331419
Panel A	1	**0.0007**	**0.0004**	-0.0000	**0.0027**	0.0003
		0.0140	**0.0740**	0.9900	**0.0620**	0.5396
		0.0000	**0.0076**	**0.0024**	**0.0004**	0.5396
	2	**0.0007**	**0.0006**	**0.0020**	**0.0043**	-0.0004
		0.0080	**0.0010**	**0.0110**	**0.0060**	0.4955
		0.0000	**0.0029**	**0.0693**	**0.0002**	0.4955
Panel B	60-Day	**0.0014**	**0.0009**	**0.0269**	**0.0073**	-0.0001
		0.0000	**0.0010**	**0.0020**	**0.0010**	0.9124
		0.0000	**0.0007**	**0.0934**	**0.0000**	0.9124
	90-Day	**0.0022**	**0.0013**	**0.0030**	**0.0110**	0.0003
		0.0000	**0.0000**	**0.0040**	**0.0000**	0.7058
		0.0000	**0.0001**	**0.0118**	**0.0000**	0.7058
	120-Day	**0.0030**	**0.0014**	**0.0023**	**0.0152**	0.0004
		0.0000	**0.0000**	**0.0060**	**0.0000**	0.5956
		0.0000	**0.0000**	**0.0150**	**0.0000**	0.5956

Table 3

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, ΔCUR_{t+j} is the currency return in month $t+j$, in panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t. The first number in each cell reports the R^2 of the regression and the second number reports the autocorrelation and heteroskedasticity-adjusted p-value of a Wald test on the null hypothesis of joint significance, H_0: $\alpha_l = 0$, $l > 0$. 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).

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

Table 4

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

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

In panel A, ΔCUR_{t+j} is the currency return in month $t+j$, in panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t. The first number in each cell reports the R^2 of the regression and the second number reports the autocorrelation and heteroskedasticity-adjusted p-value of a Wald test on the null hypothesis of joint significance, H_0: $\alpha_l = 0$, $l > 0$. 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).

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

Table 5
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

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

In panel A, ΔCUR_{t+j} is the currency return in month $t+j$, in panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t. The first number in each cell reports the R^2 of the regression and the second number reports the autocorrelation and heteroskedasticity-adjusted p-value of a Wald test on the null hypothesis of joint significance, H_0: $\alpha_l = 0$, $l > 0$. 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).

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

Table 6

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, ΔCUR_{t+j} is the currency return in month $t+j$, in panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t. The first number in each cell reports the R^2 of the regression and the second number reports the autocorrelation and heteroskedasticity-adjusted p-value of a Wald test on the null hypothesis of joint significance, H_0: $\alpha_l = 0$, $l > 0$. 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).

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

Table 7
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, ΔCUR_{t+j} is the currency return in month $t+j$, in panel B it is the excess currency return in month $t+j$. MPI_t is the estimated measure of order flow driven by marketwide private information in month t. The first number in each cell reports the R^2 of the regression and the second number reports the autocorrelation and heteroskedasticity-adjusted p-value of a Wald test on the null hypothesis of joint significance, H_0: $\alpha_l=0$, $l>0$. 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).

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