Profiting from Mimicking Strategies in Non-Anonymous Markets

Vasios, Michalis and Payne, Richard and Nolte, Ingmar

2015

Profiting from Mimicking Strategies in Non-Anonymous Markets*

Ingmar Nolte[†] Richard Payne[‡] Michalis Vasios[§]

May 24, 2013

Abstract

We explore the information content of counterparty identities and how their disclosure can be exploited by other investors in a post-trade transparent market. Using data from the Helsinki Stock Exchange, we form dynamic mean-variance strategies with daily rebalancing which condition on the net flow of individual brokers. We find that investors can benefit greatly, up to 36% in annualized risk adjusted returns, from knowing who has been trading. We demonstrate a link between the information content of broker order flow and the sophistication of their clients. Brokers who have clients that trade with a momentum style or who are predominantly institutions or foreign investors have much more informative flow than do others. In the Finnish setting, this means that brokers with large market share have uninformative flows.

Keywords: Market Transparency; Asset Allocation; Broker Heterogeneity; Customer Order Flow.

JEL Classification: C53; G11; G14; G18; G24.

*We would like to thank an anonymous referee, Nicholas Barberis, Pasquale Della Corte, Arie Gozluklu, Charles Jones, Heracles Polemarchakis, Gideon Saar, and seminar participants at Warwick Business School, the 2011 TADC Conference (London Business School), the 2011 Warwick Frontiers of Finance Conference, the 2011 High Frequency Research Training Workshop (The Marie Curie Risk Network and Manchester Business School), the 12th Symposium on Finance, Banking, and Insurance (Karlsruhe Institute of Technology), the 2012 SOFIE (Oxford) conference, the 2012 EFMA (Barcelona) conference, and the 2012 EEA-ESEM (Malaga) for helpful comments.

[†]Warwick Business School, Finance Group, University of Warwick, CV4 7AL, Coventry, UK. Phone +44 2476 572 838, Fax -23779, email: Ingmar.Nolte@wbs.ac.uk.

[‡]Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, UK. Phone +44 207 040 4114, email: richard.payne@city.ac.uk.

[§]Corresponding author: Warwick Business School, Finance Group, University of Warwick, CV4 7AL, Coventry, UK. Phone +44 7942 468 063, email: michalis.vasios.09@mail.wbs.ac.uk.

1 Introduction

The implications of the disclosure of identities of traders and those submitting orders for the outcomes (i.e. prices, trading activity and liquidity) in equity markets is an area of ongoing debate. To date, though, there is little direct empirical evidence on the information content of counterparty identities and how their disclosure can be exploited by other investors. This scarcity is surprising given that it is received wisdom in finance that, in fully transparent markets, costless observation of identities may allow investors to make a trading profit by mimicking other, better informed investors. The 'mimicking assumption' plays a key role in the literature that examines the impact of anonymity on liquidity and market efficiency. Indeed, a number of empirical studies rely on it to explain the finding that liquidity improves when anonymity is introduced.[1]

However, the validity of the mimicking argument is questionable for at least two reasons. First, as a matter of market reality, identities of traders are available at best at the broker level. This means that underlying informed client trades are aggregated with trades of uninformed customers so that the overall information content of a broker's flow is unclear. Second, using simple efficient markets logic, in a setting with post-trade disclosure of identities, one would expect that any economically significant information that could be inferred from identities should already be reflected in prices.

Our aim in this paper is to provide some direct analysis of the economic value of observing broker identities and to explore when and why they are most valuable. We form dynamic mean-variance portfolios with daily rebalancing in the fashion of Fleming et al. (2001), using daily order flows from individual brokers to forecast the cross-section of stock returns. Order flow is measured as aggressive buy trades less sell trades in an interval as in Chordia et al. (2002) and Evans and Lyons (2002).[2] We test directly

[1] See e.g., Comerton-Forde et al. (2005), Foucault et al. (2007), Comerton-Forde and Tang (2009), Friederich and Payne (2011) among others.

[2] Kyle (1985) and Glosten and Milgrom (1985) provide the theoretical foundations for the information content of order flow. Empirically, evidence from Hasbrouck (1991), Evans and Lyons (2002), Payne

whether mimicking strategies generate significant economic gains using data from the Helsinki Stock Exchange (HEX), a post-trade transparent market, in which the broker identity is publicly available. We then compare performance of the trading rule across brokers and analyze which type of broker is most profitable to mimic in order to shed light on the economics that underly our results.

We start by comparing the performance of the broker portfolios against a benchmark portfolio that disregards identity information. We show that broker portfolios outperform the benchmark portfolio. An investor with mean-variance preferences can improve his portfolio performance by up to 36% in annualized risk-adjusted terms using broker identity information. It is worth noting that if we extend the broker-level order flow measure used to include trades in which a broker was the passive counterparty, then there is no significant improvement in trading rule performance.[3] Regardless, our baseline result is that using publicly available order flow information from brokers can generate positive risk-adjusted returns.

A second result from our work is that the information content of flows varies across brokers and we dig into the reasons for this. Intuitively, one would expect that the degree of sophistication of a broker's investor base should play the key role in determining the information content of the broker's net flow. As we do not have direct access to information on broker clienteles we use proxies based on publicly available trading data for the sophistication of the broker clientele.[4]

(2003), Love and Payne (2008), and Rime et al. (2010) among others, demonstrates the important role of *order flow measure* on the transmission of private and public information into prices.

[3]Some recent work, for example Latza and Payne (2013), shows that intra-day flows of limit orders are better forecasters of stock returns than are market order flows. In this context, our results might be seen as surprising. It is worth noting, though, that the sampling frequencies of the current work and Latza and Payne (2013) differ greatly and that they use data on all orders, executed and unexecuted. In our work we have access only to completed trades i.e. executed orders.

[4]As the trade data we use to proxy sophistication is public, it can be used by any market participant to identify the most informative broker identities. This approach is in contrast to the literature on

We first examine brokers with respect to their market share. Previous literature suggests that market leaders have better information than other traders.[5] Arguments to support these results usually rely on underlying customers being more sophisticated. We test if this assertion holds in HEX, and find that a portfolio based on large broker flows significantly underperforms a portfolio based on small broker flows. The order flow information of large brokers has on average no predictive power at the daily horizon. This result is striking in that it runs counter to other results in the literature.

One explanation for this striking result, however, is that the largest brokers' net trades do not reflect the trades of sophisticated investors at all. It is possible that market leaders have very heterogeneous customer bases such that, when trades are aggregated, the noise trades from the uninformed drown the signal from the smart customers. Using the framework of Lakonishok et al. (1992) we provide empirical evidence that supports this explanation: large brokers' order flow is relatively balanced on buy and sell sides, while the trades of small brokers tend to locate either on the buy or on the sell side.

Although our result seems to run counter to conventional belief, it is consistent with prior evidence from Finland. For example, Linnainmaa and Saar (2012) report that large brokers have customers drawn fairly equally from all segments of the investor universe, but with households being their major pool of customers. On the other hand, small brokers represent mostly foreign investors and domestic institutions. As Grinblatt and Keloharju (2000) note, in Finland domestic households are the least sophisticated investors, while foreign investors are the most sophisticated.[6] Therefore it seems that the order flow of

information asymmetries across different types of investors (e.g., institutional vs. retail investor (e.g., Barber et al. (2009a)), banks vs. retail traders (e.g., Nolte and Voev (2011)) etc), as the type of an ultimate investor is virtually never made public.

[5] See e.g., Goodhart (1988), Lyons (1997), and Peiers (1997). One exception is Sapp (2002) who finds the opposite result.

[6] We quote from Grinblatt and Keloharju (2000): "All of the Finnish investor categories are probably less sophisticated than the foreign investors. Foreign investors tend to be well capitalized foreign financial institutions with a long history of successful investment in other stock markets. This category is generally

large brokers is likely to be an aggregation of noise, while small brokers aggregate the trades of better informed and more sophisticated investors.

In a second step, we examine whether brokers who have unusually large order flows (regardless their average market share) on a particular day-stock convey more information to market than do others. We regard this exercise as an attempt to identify higher frequency (i.e. daily) variations in information content of brokers' net trades. The intuition is simple. In a post-trade transparent setting, earlier work (e.g., Rindi (2008)) suggests that sophisticated traders will exploit their information advantage very aggressively in an attempt to mitigate problems from information leakage[7]. Our analysis supports this view. The order flow of the most active brokers on a particular day, has a major impact on the following day's prices. This impact lasts for one day only, however. Hence although large average trading volumes are rather uninformative, trading concentration in a short period of time does capture 'smart trading activity'.

Another sophistication proxy identified by prior literature is the use of momentum trading styles. Both Grinblatt and Keloharju (2000) and Linnainmaa (2010) show that the more sophisticated investors in Finland exhibit a momentum investment style. In other markets, Grinblatt et al. (1995) and Goetzmann and Massa (2002) find similar patterns. Using the framework of Grinblatt and Keloharju (2000), we characterize the trades of each broker as being either momentum or contrarian on average. Then, we show that the order flow of brokers whose customers trade with a momentum style in aggregate (momentum brokers) better predicts future returns than flow from brokers with contrarian clients. Again, sophistication is proxied using observable data and brokers with more sophisticated clienteles have more informative flows. We provide some further evidence to support the link between brokers whose trading activity exhibits momentum and sophistication as we measure each broker's stock picking ability and show that momentum brokers are better

composed of mutual funds, hedge funds, and foreign investment banks".

[7]Another strategy to hide information could by the usage of multiple brokers. However, evidence from Linnainmaa and Saar (2012) suggests that trading through multiple brokers is not common in Finland.

stock pickers.

Overall our results support the notion that observing broker level order flow, even when it is publicly available, can allow one to earn positive excess returns. The result that identity information is valuable is in line with expectations, but it is worth noting that all of the information in our portfolio construction experiments is public such that the results provide a challenge to full market efficiency. More strikingly, we show that public data can also be used to discriminate between brokers and to identify those most likely to have informative flows. Brokers who have clients that trade with a momentum style or who are predominantly institutions or foreign investors have much more informative flow than do others. In the Finnish setting, this means that brokers with large market share have uninformative flows. Our work clearly has elements in common with Linnainmaa and Saar (2012), who also look at Finnish data and who categorise brokers by the structure of their customer base. While they use proprietary data to make this link, we show that one can roughly approximate client sophistication using data that is publicly distributed. Moreover, we show that this information can be exploited to generate economic value through the use of a simple trading rule.[8]

Given that our analysis does not study a change in anonymity regime, our results can say little about the direct linkages between liquidity, efficiency and transparency. However, many of the models which do link transparency to market quality rely on the mimicking assumption that we test in this paper. To the extent that we find evidence that mimicking the trades of more sophisticated agents can lead to economic gains, we support the notion that full transparency might be costly for markets, and thus investors. It may discourage sophisticated investors from expending effort to uncover information leading to less efficient prices and lower liquidity. It should be noted, though, that the

[8]It is also worth noting that Finnish equity market microstructure changed in between our sample and that of Linnainmaa and Saar (2012). In their sample, the market was pre and post trade non-anonymous, while in our data sample the market was only pre-trade anonymous. This means that in our sample, one could only gain information about trader identities from trade reports, as we do.

fact that we find evidence of profits from mimicking implies that this strategy is not being so aggressively pursued by market participants that the opportunity disappears.[9]

The rest of the paper is organized as follows. The next section describes the data set we use. Section 3 presents the empirical framework. In Section 4, we discuss the empirical results of the baseline mean-variance analysis. In Section 5, we investigate the determinants of the information content of broker order flow and broker heterogeneity. Section 6 concludes.

2 Data and Summary Statistics

We use intraday equity data from the Helsinki Stock Exchange (HEX), which are provided by Bloomberg. HEX has been part of NASDAQ OMX Group since 2007. As advertised on their website, the NASDAQ OMX Group is the world's largest exchange company. NASDAQ OMX Nordic describes the common offering from NASDAQ OMX exchanges in Helsinki, Copenhagen, Stockholm, Iceland, Tallinn, Riga, and Vilnius. These exchanges use a common trading platform, which allows for cross-border trading and settlement, and cross membership.

Our data set begins at 8am (GMT) on Monday 29^{th} March 2010 and ends at 4:30pm on Monday 28^{th} February 2011; this amounts to 210 trading days. In our analysis, we consider the 15 most liquid (in terms of turnover) stocks of the HEX25 index, in order to circumvent problems arising from the low number of transactions of some brokers. Table 2 reports the summary statistics of the 15 stocks. Every day there are 4 regular trading sessions: opening (7am-8am), continuous trading (8am-4:25pm), closing (4:25pm-4:30pm) and after market (4:30pm-7am). We restrict the empirical analysis to the continuous trading session. HEX is classified as a post-trade transparent market, as in addition to the typical information on price and volume, the identity of the counterparty who bought

[9]Interested readers might read a related policy discussion on the implications of introducing transparency in dark pools on the "Securities and Exchange Commission; Concept Regulation of Non-Public Trading Interest; Proposed Rule, November 23, 2009, Federal Register 74(224), 61207-61238.

and sold is also provided for executed trades. The counterparty identity is not available for limit orders - pre-trade anonymity. The investor type behind each trade, such as whether it is an end-investor or a trade on broker's own account is also not provided.

The raw data contains information on 7 items, as shown in the 5 second span of Nokia transaction data shown in Table 3. The first two columns are the date and time expressed as month/day/year and hour:minute:second, respectively. The third column is the type of the transaction, which can be "Best Bid", "Best Ask" or "Trade". The next two columns are the price (in Euros) and the size of the transaction. The last two columns are the Broker Buy Code and the Broker Sell Code. In Table 1 we list the brokers from HEX.[10] The counterparty identity is available only for transactions of type "Trade".

The data record does not provide the direction of trade. However, the availability of best bid and best ask quotes, as well as their time stamp enables us to identify which broker initiated the trade and, thus, the direction of the trade. This identification is an important element in our empirical analysis, since it allows us to disaggregate the data and construct distinct order flow measures for every broker.

3 Empirical Framework

3.1 The Formation of Dynamic Mean-Variance Portfolios

Our empirical analysis relies on the formation of dynamic mean-variance portfolios. Our investment scenario considers an investor with mean-variance preferences, who allocates his wealth across the 15 most liquid stocks of the HEX and the risk-free asset. Rebalancing

[10] We drop brokers that do not trade in all stocks, are acquired by other brokers, are not members for the entire sample period, and those who initiate (on average, across stocks) less that 10 trades every day and are active (on average, across stocks) in less than half of the days of our sample period. We do that to deal with the computational problems arising from zero observations when we estimate the order flow models (Section 3), and to increase the power of the LSV statistic and the buy ratios (Section 5). The results remain qualitatively the same if instead we use all brokers.

is daily and conditional on the observation of the previous day's trading activity, which is captured by order flow measures. HEX being a non-anonymous market allows investors to observe not only the aggregate market order flow, but also the customer order flow of brokers. Either order flow is an input to investor's optimization problem, which is given by:

$$
\begin{aligned}
\max_{w_t^j} \ \mu_{p,t+1|t} &= w_t^{j\,\prime}\mu^j_{N,t+1|t} + (\iota - w_t^{j\,\prime}\iota)R_t^f \\
s.t. \ \ \sigma_p^2 &= w_t^{j\,\prime}\Sigma_{t+1|t}w_t^j, \qquad t = 1,\ldots,T,
\end{aligned}
\tag{1}
$$

where $j = 1,\ldots,J$ identifies broker^j, w_t^j is the $N \times 1$ vector of portfolio weights; $\mu_{p,t+1|t}$ is the conditional expected portfolio return; σ_p is the target portfolio volatility; $\Sigma_{t+1|t}$ is the $N \times N$ variance-covariance matrix of the risky assets; R_t^f is the risk free rate; and $\mu^j_{N,t+1|t}$ is the $N \times 1$ vector of expected returns of the risky assets conditional on the order flow information of broker^j, $\mu^j_{N,t+1|t} = E[R_{t+1}|\Im_t^j]$. The solution to this constrained maximization problem yields,

$$
w_t^j = \frac{\sigma_p \ \Sigma^{-1}_{t+1|t} \ (\mu^j_{N,t+1|t} - \iota R_t^f)}{\sqrt{(\mu^j_{N,t+1|t} - \iota R_t^f)' \ \Sigma^{-1}_{t+1|t} \ (\mu^j_{N,t+1|t} - \iota R_t^f)}}. \tag{2}
$$

These are the weights for the risky assets at each rebalancing time interval. The investment in the risk free asset is equal to $1 - w_t^{j\,\prime}\,\iota$. Then, the period $t+1$ gross return on the investor's portfolio is given by $1 + w_t^{j\,\prime}R_{t+1} + (1 - w_t^{j\,\prime}\iota)R_f$.

A key element in Equation 2 is the vector of conditional expected returns of risky assets. We presume that the information set of the aggregate market and that of brokers differ. We approximate these information sets by using transaction data to compute order flow measures. There is an extensive literature on order flow and how it can impact returns not only through short-term liquidity and inventory effects, but also because it conveys information[11]. Our methodological contribution to this literature is the disaggregation of

[11]See e.g., Hasbrouck (1991), Chordia et al. (2002), Evans and Lyons (2002), Easley et al. (2002), Payne (2003), Pasquariello and Vega (2007), Evans and Lyons (2008), Berger et al. (2008), Love and Payne

the order flow measure at the aggregate market and the broker firm level. That means that for every stock, we have as many conditional expected return estimates as the number of brokers plus the aggregate market estimate.

In our analysis, we use 2 order flow specifications. The first one is the standard *order flow* (OF_t^j) measure of Chordia et al. (2002) and Evans and Lyons (2002), defined as the daily buyer-initiated volume minus the seller-initiated volume. This measure captures aggressive trading, which is considered to transmit new information into prices. To see whether liquidity-supplying (passive trading) conveys information too, we use a second specification, the $VolOF_t^j$, defined as the total buy volume minus the total sell volume executed by broker^j[12].

Building on these order flow measures, we use 2 parsimonious models to compute one-day-ahead estimates of stock returns[13]. The first model (M1) is a pure order flow model:

$$R_{t+1}^i = \alpha + \beta OF_t^{ij} + \epsilon_{t+1}, \qquad i = 1, \ldots, 15, \tag{3}$$

where j identifies broker^j, R_{t+1}^i is the return of stock i, OF_t^{ij} is the order flow measure of broker^j on stock i, β is a coefficient, α is a constant, and ϵ_{t+1} the error term. To capture liquidity supply, Model 2 (M2) uses the second order flow measure, $VolOF_t^{ji}$, as an additional variable:

$$R_{t+1}^i = \alpha + \beta OF_t^{ij} + \gamma VolOF_t^{ij} + \epsilon_{t+1}, \tag{4}$$

(2008), Nolte and Nolte (2010).

[12]To clarify things, the difference between OF_t^j and $VolOF_t^j$ is that the former used only marketable orders, while the latter uses all trades, which includes aggressive (marketable orders) and passive (limit orders) volume. We are able to calculate $VolOF_t^j$ because our dataset contains the identity of the broker that bought and sold in every transaction. By construction, this order flow definition is zero for the aggregate market; the daily buy volume always equals the daily sell volume.

[13]Rime et al. (2010) use order flow measures to predict returns in a foreign exchange context.

In unreported results we also include the market return, the HEX25 index, in an attempt to capture market-wise economic activity at time t and control for momentum effects. However, we find that market return does not play any significant role.

3.2 Performance Measures

The next step in our analysis is to measure the performance of the mean-variance portfolios. We use an economic evaluation approach and two criteria; a traditional performance measure, the Sharpe ratio, and a utility-based measure, the manipulation-proof performance measure (MPPM) of Goetzmann et al. (2007). The first economic criterion, the ex-post Sharpe ratio (SR), is defined as:

$$SR^j = \frac{\overline{R_p^j - R_f}}{\sigma_p^j}, \tag{5}$$

where the nominator is the average (annualized) excess portfolio return and the denominator is the portfolio's (annualized) standard deviation. Intuitively, the Sharpe ratio measures the risk-adjusted annualized portfolio's returns.

The second economic criterion, MPPM, is defined as:

$$MPPM^j = \frac{1}{(1-\gamma)\Delta t} \ln \left[\frac{1}{(T-1)} \sum_{t=1}^{T-1} \left(\frac{R_{p,t+1}^j}{R_{t+1}^f} \right)^{1-\gamma} \right], \tag{6}$$

where $R_{p,t+1}^j$ is the gross portfolio return obtained when using broker's j order flow to forecast expected returns, R_{t+1}^f is the gross risk free return, Δt is the one day interval, and γ can be seen as the investor's relative risk aversion coefficient. $MPPM^j$ can be interpreted as the annualized continuously compounded excess return certainty equivalent of the portfolio that uses broker^j order flow information to predict returns. The advantage of this economic measure is that it does not require an assumption of the investor's utility function, and it is robust to the distribution of the portfolio returns.

Our interest lies on the performance differences rather than on the performance of the mean-variance portfolios per se. We therefore use as a benchmark the portfolio that uses the aggregate market order flow (ANON), which is the one that disregards the broker identity. The performance difference against the ANON portfolio allows us to measure the predictive power of the customer order flow of brokers. If broker identity contains no information, this difference should be zero. In contrast, a positive performance difference will unveil the predictive power of broker identity[14]. The latter is not an obvious outcome. One reason is that brokers "collect" and execute orders most probable from a diverse pool of investors, which contains not only informed, but also uninformed traders. In addition, even in the case of informed trading, the disclosure of identities takes place post-trade, which means that prices might already reflect any valuable information. Therefore the predictive power of broker identity is an empirical question.

We test this hypothesis by calculating the following performance difference measure:

$$\Theta^j = MPPM^j - MPPM^{ANON}. \tag{7}$$

Θ^j enables us to compare competitive dynamic investment strategies. Intuitively, it is the fee that a mean-variance investor is willing to pay to switch from the benchmark asset allocation strategy to the strategy under investigation. A positive Θ^j will mean that the investor will be better-off using broker^j order flow information than using the aggregate market order flow information.

There are a number of papers that use utility-based measures to determine the economic value of a dynamic strategy versus a passive strategy. For instance, Fleming et al. (2001) investigate the economic value of volatility timing and Marquering and Verbeek (2004) analyze the economic value of predicting both stock index returns and volatility.

[14]Intuitively, in a market where some investors have privileged access to customer order flow (e.g. in a anonymous market in which brokers reveal their customer order flow to their favorite investors) the value of knowing who trades will be even larger.

While we follow a similar approach, a critical difference is the fact that our benchmark strategy is not passive, but dynamic. Here, the information sets captured by our order flow models differ, not the style of investment strategy. By selecting the same dynamic strategy for the benchmark portfolio, we can isolate the effect of market transparency on portfolio performance from other effects associated with the style of asset management (i.e., active versus passive management)[15].

4 Empirical Results

In this section, we measure the predictive power of flows form individual brokers. The investment scenario is based on an investor with a coefficient of relative risk aversion of 6, who maximizes his expected portfolio return subject to an annual target volatility of σ_p=10%. Our choice of σ_p and γ is consistent with previous literature (see e.g., Rime et al. (2010) and Della Corte et al. (2010)). Choosing alternative values of σ_p and γ leaves our results qualitatively unchanged. Our choice of risk free rate is the one month eurodollar rate (mnemonic is ECEUR1M), which is available on a daily basis. We use order flow models M1-M2, described in Section 3.1 to predict returns and rebalance portfolio weights on a daily basis. This recursive out-of-sample regression estimation is based on a window of expanding size that means that the investor uses all available historical information on day t to update his beliefs and optimize his asset allocation on day $t+1$. The initial estimation window is 03/30/2010–08/09/2010 (86 days or 40% of the sample period) and the portfolio formation and rebalancing runs from 08/10/2010 to 02/28/2011 (124 days or 60% of our sample period).[16] We compute the variance-covariance matrix of the risky assets recursively on each day using data from the previous one year to forecast volatility at t+1.

[15] Our results do not change qualitatively when a passive benchmark strategy is chosen instead.

[16] Results remain qualitatively similar if instead we split the sample period in two equal windows.

4.1 Preliminary Analysis

Before we proceed to the recursive out-of-sample estimation we present some preliminary results to obtain an indication of the statistical performance of the order flow models in the initial estimation window. To save space we present results only for model M1. Results for the other model are available upon request.

Figure 1 gives a picture of the heteroskedasticity and autocorrelation corrected (Newey-West) t-statistics on the lagged order flow variable (OF) in model M1. M1 is the model in Equation 3, in which we regress the returns of stock i (R^i) on the previous period's order flow measure, which is calculated either at the aggregate market level (i.e. OF^{iANON}) or at the individual broker level (i.e., OF^{ij}). The first statistic (blue bars) shows the number of positive order flow coefficients, while the second statistic (red bars) is the number of statistically significant coefficients. As we can see, on the one end, there are brokers like SHB or SWB with many significant coefficients, while at the other end there are brokers like NIP and NON with barely any significant coefficients. The sign of the order flow coefficient is positive for half of the estimates, while for almost one third of the brokers, the majority of their coefficients is positive. A positive coefficient indicates that order flow and next day's return are positively related; buy (sell) pressure on a particular trading day predicts a price increase (decrease) the next trading day, although it remains to be sees whether this is reflected in the portfolio performance.

In Table 4, we move from statistical to economic evaluation. We want to stress that results in this table are from an in-sample estimation, since the order flow coefficients are estimated only once using the first 86 days of the dataset and portfolios are constructed for the same 86 days. In the out-of-sample estimation, we repeat this computation in each of the next 128 days. In short, the results show that investors can improve their portfolio performance when they observe brokers' customer order flow compared to the benchmark case, which is the portfolio that disregards the broker identity. The number of positive Θ ranges from 18 for Model 1 to 22 for Model 2. However, not all of these Θ

are statistical significant, with the best model being Model 2 with 6 statistical significant Θ (p-value$<$ 10%)[17]. Among the brokers that perform well across all models are DBL, DDB, JPM, NRD, and SWB. As for the Sharpe ratios, they are high across all brokers and models, which is expected as these are in-sample calculations with daily rebalancing. Their magnitude is consistent with other papers that use order flow models with daily rebalancing. For instance, Rime et al. (2010) find in-sample Sharpe ratios that range from 5.79 to 7.05.

The results in this section support our hypothesis regarding the positive economic value of market transparency. However, the real test lies in the out-of-sample evaluation of the recursive forecasts and the performance of the mean-variance portfolios that follow in the next section.

4.2 Does Broker Identity Convey Information?

In this section, we test the hypothesis that the broker identity conveys information and that investors can benefit from transparency. Our analysis is based on an out-of-sample recursive regression estimation.

Table 5 presents the economic evaluation of the mean-variance portfolios. For the majority of brokers, Sharpe ratios are large, positive and greater than the Sharpe ratio of the ANON portfolio, which is negative. The same holds for Θs. More specifically, the reported p-values are below at least 10% (5%) for 11 (7) brokers for Model 1 and, 7 (2) brokers for Model 2. To better illustrate the results, in Figure 2 we present the brokers in descending order with respect to Θ. It is clear that on average broker portfolios outperform the benchmark portfolio. The Θs are positive and significant for almost one third of brokers. Among the best performer brokers are CAR, NRD, RBN, SHB and SWB. One interpretation of a positive Θ is that it measures the maximum performance

[17]The number of significant Θ increases to 9 when the market portfolio is included, which controls for market-wise developments and momentum effects. For brevity, we do not report these results.

fee the mean-variance investor is willing to pay to switch from the ANON portfolio, which is the one that disregards the broker identity, to the portfolio that tracks broker^j customer order flow. Hence we conclude that the broker identity conveys economically significant information.

Another finding in Table 5 is that the predictive power of broker identities mainly comes from the marketable limit orders, which captures aggressive trading. In contrast, passive trading activity, $VolOF_t^j$, most often decreases the magnitude and significance of the Θs. The trades that brokers initiate are more informative than those they participate in but which are initiated by others.

In the literature evaluating the performance of dynamic investment strategies, transaction costs play a key role. However, in unreported results, we find that the predictive power of identities is robust to the presence of transaction costs; transaction costs either play a minor or even a supportive role. This is in line with our expectations, as in our framework the benchmark portfolio follows also a dynamic investment strategy, thus, in the presence of transaction costs its performance will be affected too. These results are available upon request.

The evidence to this point suggests that even in a post-trade transparent market the information about broker identities is useful for other investors: market transparency yields positive economic value. Specifically, investors can erase the noise from the aggregate market by observing brokers' customer order flow and greatly improve their investment decision making up to 36% (annualized) percentage points (Model 1). This finding is consistent with the theoretical models by Forster and George (1992) and Benveniste et al. (1992), that suggest that the observation of identities can give rise to private benefits and trading profits. Other models that examines the effects of dual trading and front running (e.g., Roell (1990), Fishman and Longstaff (1992)) also suggest that observing who trades can produce trading profits. However, in these studies the observation of who trades is a privilege of only a few market participants, while in our setting the information about the trades of brokers is public information.

Another finding from Table 5 is the strong heterogeneity in results across brokers. Sharpe ratios range from -1.61 to 2.93 for Model 1. Θ varies too: from -10 to 36 and -15 to 38 for Models 1 and 2, respectively. It is not immediately clear why the predictive power of identity is strong for some brokers and zero for some others. From a practical point of view, it is important to know what drives this heterogeneity in order to understand the dynamics of information generation and aggregation at the broker level. Intuitively, one would expect that the degree of sophistication of brokers' investor base should play a key role. We elaborate upon this issue in the next section.

5 The Determinants of the Information Content of Broker Customer Order Flow

Our analysis suggests that the dissemination of data on who is trading and in which direction can help others investors to make better investment decisions. However, this result depends on the ex-ante ability of investors to select the brokers with the most informative customer order flow. Here, we explore the determinants of the information content of brokers' order flow. Intuitively, we expect the brokers with a more sophisticated client base to have more informative trades. Therefore we test several hypotheses, using only on publicly available information (e.g., market share and investment style), in an attempt to understand what drives the predictive power of order flow at the broker level, and whether it can be attributed to observable broker-specific characteristics.

5.1 The Role of Market Share: Large vs. Small Brokers

We start by exploring the role of brokers' market share. The simple intuition underlying the market share hypothesis is as follows. Investors pay close attention to the trading activity of market leaders. Who wouldn't take into account the trades of Goldman Sachs or other big players? Several papers, provide the reason why; banks with large market share are, on average, better informed (see e.g., Goodhart (1988), Lyons (1997), and Peiers (1997)). This is especially likely to be true in markets in which only a few brokers control most of the trading activity. As shown in Table 6, HEX belongs to this category. We calculate market share as the average daily volume initiated (only marketable limit orders) or executed (all trades) by each broker across the 15 most liquid stocks of the HEX. Table 6 shows that the brokerage industry in HEX is highly concentrated; the top 5 brokers initiate almost 40% of the trading and execute 35% of the volume.

To test whether the information content of flows varies with broker market share, we split brokers into quartiles according to our 2 volume measures, see Table 6. Then we construct daily order flow series for the top and bottom quartiles, and repeat the formation, rebalancing, and evaluation of the two mean-variance portfolios following the steps described in Section 3. We report the performance difference $\Delta\Theta$ that is defined as:

$$\Delta\Theta = \Theta_{Q4} - \Theta_{Q1}, \tag{8}$$

where Θ_{Q4} is the MPPM of top quartile and Θ_{Q1} is the MPPM of the bottom quartile. A positive and significant $\Delta\Theta$ will indicate that large brokers have more informative order flow than small brokers.

Table 7 presents the performance differences between large and small brokers, and the associated p-values. Surprisingly, we find evidence that rejects the market share hypothesis. The results in panel a. and b. show that the large broker portfolio significantly under performs the small broker portfolio. In other words, the order flow of brokers that do the largest amounts of trading, whether aggressive or passive, is less informative than the order flow of small brokers. These results suggest that the conventional belief that the order flow of large brokers conveys information is not necessarily true.

One explanation could be that in Finland ‘smart investors’ trade mainly through small brokers, while ‘naive investors’ prefer large brokers. Perhaps the clientele base of all brokers consists of all types of investors but large brokers, because of their size and reputation, attract on average the most heterogeneous customers with respect to level of sophistication, investment strategies, and ultimately beliefs. If small brokers appeal more to smart traders then their flows will be less noisy and more informative than that of large brokers.

To elaborate on this heterogeneity of investors argument, we calculate correlated trading statistics. Intuitively, there should be a negative relation between correlated trading and broker clientele diversification; the more heterogeneous the investors are, the more uncorrelated their aggregate trading activity will be. We use the Lakonishok et al. (1992) framework to explore this relation.[18]

We define the LSV statistic as:

$$H_t(j,i) = |B_t(j,i)/N_t(j,i) - p(j,t)| - AF_t(j,i), \tag{9}$$

where $B(j,i)$ is the number of broker^j trades in stock i during day t that are aggressive purchases, $N(j,i)$ is the number of all trades initiated by broker^j in stock i during day t, $p(j,t)$ is the expected proportion of all broker^j trades that are purchases on day t, and $AF(j,i)$ is an adjustment factor that captures that the first term of the formula can be greater than zero under the null hypothesis of no correlated trading. In our calculation we account for the splitting of large orders effect in the same second, which otherwise will artificially increase the LSV measure. The LSV statistic is computed for each stock-day and then averaged per broker[19].

The results in Panel a. of Table 8, report correlated trading statistics, which vary

[18] Recent papers that use the same framework are: Grinblatt et al. (1995), Wermers (1999), Barber et al. (2009a), and Barber et al. (2009b).

[19] To illustrate our approach, suppose that in a given day half of the transactions initiated by broker j are buys and half are sells. We can use this information to infer that broker j clients are trading independently, and the LSV statistic will be close to zero. On the contrary, if 90% percent of broker's j trades are buys, then we would conclude that broker j trading is highly correlated, and the LSV statistic will be greater than zero.

from 2% to 27%, for brokers CDG and RBN, respectively.[20] That means that on average the 52% of the trades initiated by broker CDG every day are on one side of the order book, buys or sells, while this number increases to 77% for RBN. The most interesting result is presented in Panel b: the average size of brokers in the bottom LSV quartile (Q1) is four times the average size of the brokers in the top LSV quartile (Q4). This result supports our argument that large brokers attract very heterogeneous customers. Panel c., shows the underperformance of the portfolio that tracks brokers with very uncorrelated trading. Clearly, our results document a connection between clientele heterogeneity, market share, and predictive power of broker identity. It suggests that it is large brokers' clients' heterogeneous trading and dispersion of beliefs that causes the rejection of the market share hypothesis. Meanwhile, it is still ambiguous what drives the strong predictive power of small brokers. The heterogeneity argument alone cannot explain this. In addition we require that small brokers are used by, on average, 'smart investors'.

As we do not have investor transaction records, we rely on prior evidence from Finland to provide supporting evidence. A recent paper of Linnainmaa and Saar (2012) utilizes trading records from Helsinki and reports that large brokers' clientele is roughly evenly split into the three major categories of investors, i.e. households, domestic institutions and foreign investors. In fact, their major pool of customers is domestic households. On the contrary, the investor base of small brokers is dominated by foreign investors, followed domestic institutions. As Grinblatt and Keloharju (2000) note, in Finland domestic households are the least sophisticated investors, while foreign investors are the most sophisticated. Specifically, they argue that "[A]ll of the Finnish investor categories are probably less sophisticated than the foreign investors. Foreign investors tend to be well capitalized foreign financial institutions with a long history of successful investment in other stock markets. This category is generally composed of mutual funds, hedge funds,

[20] These results are consistent with the previous work of Dorn et al. (2008) and Barber et al. (2009b), who document correlated trading among the clients of a German and a U.S. broker, respectively.

and foreign investment banks". Therefore it seems that our results are driven by the combination of heterogeneous client bases for large brokers and sophisticated clients for small brokers. This means that large broker flow will be at best a noisy signal of future returns while small broker flows will be much better predictors.

5.2 Active Brokers

Although the trades of large brokers are rather uninformative, in this section we explore whether brokers who have unusually large order flows (positive or negative) on a particular stock-day convey more information to market than do others. Our motivation is driven by the typical argument in the market anonymity literature (e.g., Rindi (2008)) that in a post-trade transparent setting the sophisticated traders will aggressively exploit their information in an attempt to reduce information leakage. Intuitively, their aggressive trading will be reflected on the order flow of the most active brokers, in terms of their customer trading volume, on a particular stock-day.

The key difference from the previous section is that now we proxy sophistication at a higher frequency (i.e. a day), and we allow mean-variance investors to update their beliefs about the brokers with the more informative order flow on a daily basis. We therefore form a dynamic mean-variance portfolio that tracks the most active brokers on day t and stock i in order to predict future returns of stock i on day $t+1$, $t+2$ and $t+3$. Table 9 presents the performance against the ANON portfolio and the associated p-values. We find that the customer order flow of brokers who initiate heavy volume on a particular day has very strong predictive power for the next day's returns (Panel a). The predictability for returns on day $t+2$ (Panel b) and $t+3$ (Panel c) is statistically insignificant. Hence the results of this section provide further support for the linkage between smart or sophisticated clients and informativeness of order flow.

5.3 Past-return-based Investment Style and Sophistication

Prior literature also links sophistication with investment patterns based on past returns, particularly momentum trading.[21] The typical argument is that informed and sophisticated investors exhibit an momentum investment style: they buy past winners and sell past losers. On the contrary, 'naive investors' exhibit a contrarian investment style: they buy past losers and sell past winners. In this section, we first explore whether these behavioral patterns survive at the broker aggregation level, and then whether discriminating between brokers with momentum and contrarian based flows can be useful for predicting returns.

We follow the Grinblatt and Keloharju (2000) framework to characterize brokers in terms of their investment style. This framework consists of measuring the difference between the buy ratio of past winning stocks (top quartile) and the buy ratio of past losing stocks (bottom quartile). The buy ratio of broker^j is defined as:

$$\text{Buy Ratio}^j = \frac{\text{Buy Volume}^j}{\text{Buy Volume}^j + \text{Sell Volume}^j}, \tag{10}$$

where all volumes are calculated using trades initiated by broker j. We compute daily and hourly buy ratios in order to capture both the daily and intradaily patterns. If the difference is positive (negative), then the broker is viewed as momentum (contrarian) oriented at time t. We calculate buy ratio differences for every time interval, and if the fraction of days (or hours) with positive differences is higher (lower) than 0.50, the broker displays momentum (contrarian) behavior. We analyze statistical significance with both the standard two-sided binomial test and the AR(1)-adjusted binomial test suggested in

[21] See e.g., Grinblatt et al. (1995), Goetzmann and Massa (2002), Griffin et al. (2003). Bloomfield et al. (2009) show that short-term momentum is mainly caused by sophisticated informed traders. Grinblatt and Keloharju (2000) find that in HEX the more sophisticated investors are, the more momentum is their behavior. Hvidkjaer (2006) reports evidence for informed trading among large traders, whose investment style is momentum.

Grinblatt and Keloharju (2000).[22] To save space we do not report the p-values of the second test, as they are very similar to those produced by the standard test.

Table 10 presents the fractions of positive buy ratio differences for the intraday and daily horizon, along with the p-values of the associated binomial tests. We observe strong behavioral patterns, both reversal and momentum, at all frequencies. The fraction of positive buy ratio differences varies from 0.34 to 0.60 and 0.28 to 0.68 at the 1 hour and 1 day horizon, respectively. At the daily horizon, FOR, SAB, and UBS are the brokers with the stronger momentum behavior (65%, 68%, and 59%, respectively), and AAL, DBL, and NRD the brokers with the stronger contrarian behavior (39%, 28%, and 35%, respectively). When we move to the intraday frequency, reversal patterns become stronger and the number of significant contrarian brokers doubles.

Our results are broadly consistent with findings of the previous literature. Grinblatt and Keloharju (2000) show that investors in Finland exhibit both contrarian and momentum behavior at daily horizon, depending on their degree of sophistication, with the least sophisticated investors being contrarian. Linnainmaa (2010) also documents reversal effects using data from HEX. More recently, Heston et al. (2010) find that the strong intraday return reversals in NYSE are reversed at the daily frequency, a finding that resembles the weakening of contrarian behavior at the daily frequency in our sample.

Next, we measure the performance difference between a portfolio that tracks flows from momentum brokers and a portfolio that uses contrarian broker flows. We apply the analysis of the previous section (see Equation 8) and test whether the customer order flow of the statistically significant (p-value$<$ 5%) momentum brokers contains bet-

[22] The z-test statistic of this test is defined as:

$$z = \frac{x - T/2}{\sqrt{T/4 + [(2p-1)^{T+1} - T(2p-1)^2 + (2p-1)(T-1)]/16(1-T)^2]}}, \qquad (11)$$

where x is the the fraction of positive buy ratio differences, p is the observed proportion of continuations, and T is number of trading days.

ter information than that of the statistically significant contrarian brokers. We use the characterisations of momentum and contrarian behaviour based on daily data. Table 11 presents the results for Models 1 to 2. We find that the order flow of brokers whose average customer exhibits a momentum investment style has statistically strong predictive power for future returns. In contrast, the order flow of contrarian brokers has zero predictive power. As for the performance differences, $\Delta\Theta$, they are positive, varying from 9 to 28 percentage points. We therefore conclude that the order flow of momentum brokers is more informative than the order flow of contrarian brokers.

The evidence that the order flow of momentum investors conveys information is in line with the earlier literature that find a linkage of this particular trading style and the sophistication level.[23] Grinblatt and Keloharju (2000) suggest that another way to capture this linkage is by constructing measures of stock picking ability, which is what we do next. In particular, we measure the stock picking ability of brokers by examining the buy ratios of future returns. If the average buy ratio of future winning stocks (top quartile) exceeds the buy ratio of future losing stocks (bottom quartile), then this provides evidence of high stock picking ability and, thus, evidence of sophistication. Future returns are the cumulative daily returns of the next 1 month and 3 months. We compute buy ratio differences for every day and if the fraction of days with positive differences is higher (lower) than 0.50, the broker displays high (low) stock picking ability.

We focus on the two extreme quartiles - the brokers with the highest stock picking ability (Q4) and the brokers with the lowest stock picking ability (Q1) - and compare the average investment style of each group. Table 12 shows that the differences of the two groups are large and statistically significant from zero (p-value$\approx$ 0%): 0.57 vs. 0.48 and 0.56 vs. 0.46 for the 1 month and 3 months horizon, respectively. Hence, in line with our

[23] In unreported results, we examine if the returns of the momentum brokers' portfolio can be explained solely due to a momentum premium. We find that the momentum premium cannot fully explain the momentum brokers' portfolio returns, leaving space for the superiority of their customer order flow information story.

expectations we find additional supportive evidence to the linkage of the brokers' client sophistication with momentum behavior.

6 Conclusion

The financial crisis of 2008 has triggered a discussion in the finance community about the need to move to more transparent market structures. Policy makers in the US and Europe have already taken steps into this direction. One of the dimensions of transparency that is in play concerns the disclosure of counterparty identity. Yet there is little direct empirical evidence on the information content of counterparty identities in brokered markets and on how their disclosure might impact other investors, especially in the post-trade disclosure case.

In this paper, we provide direct empirical evidence on these issues. In particular, we explore whether counterparty identities convey information and, if so, how this information can be exploited by other investors. To facilitate our analysis we use data from Helsinki Stock Exchange, in which broker identities are (post-trade) publicly available. Within a simple mean-variance framework with daily rebalancing we show that broker identity conveys economically significant information. More specifically, we find that mean-variance investors can improve their trading profits by observing the order flow of individual brokers. This is translated into a superior portfolio performance up to 36% annualized percentage points compared to the benchmark investment scenario, in which investors disregard broker identity information.

We next investigate why the identity of some brokers is more informative than others and find that there is a linkage between the information content of their order flow and the sophistication of their client base. It is worth noting that we use only publicly available transaction data to proxy for client sophistication. Our analysis shows that brokers who have clients that trade with a momentum style or who are predominantly institutions or foreign investors have much more informative flow than brokers with a more contrarian and heterogeneous client base. Surprisingly, in the Finnish setting, this means that brokers with large market share have less informative flows. Overall, our work suggests that the disclosure of identities can lead to the existence of profitable mimicking strategies, and investors can earn positive excess returns by utilizing public identity information.

References

Barber, B., Odean, T., Zhu, N., 2009a. Do retail trades move markets? Review of Financial Studies 22, 151.

Barber, B., Odean, T., Zhu, N., 2009b. Systematic noise. Journal of Financial Markets 12, 547–569.

Benveniste, L., Marcus, A., Wilhelm, W., 1992. What's special about the specialist? Journal of Financial Economics 32, 61–86.

Berger, D., Chaboud, A., Chernenko, S., Howorka, E., Wright, J., 2008. Order flow and exchange rate dynamics in electronic brokerage system data. Journal of International Economics 75, 93–109.

Bloomfield, R., Tayler, W., Zhou, F., 2009. Momentum, reversal, and uninformed traders in laboratory markets. Journal of Finance 64, 2535–2558.

Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order imbalance, liquidity, and market returns. Journal of Financial Economics 65, 111–130.

Della Corte, P., Sarno, L., Sestieri, G., 2010. The predictive information content of external imbalances for exchange rate returns: How much is it worth? Review of Economics and Statistics , forthcoming.

Dorn, D., Huberman, G., Sengmueller, P., 2008. Correlated trading and returns. Journal of Finance 63, 885–920.

Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? Journal of Finance 57, 2185–2221.

Evans, M., Lyons, R., 2002. Order flow and exchange rate dynamics. Journal of Political Economy 110, 170–180.

Evans, M., Lyons, R., 2008. How is macro news transmitted to exchange rates? Journal of Financial Economics 88, 26–50.

Fishman, M., Longstaff, F., 1992. Dual trading in futures markets. Journal of Finance 47, 643–671.

Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. Journal of Finance 56, 329–352.

Forster, M., George, T., 1992. Anonymity in securities markets. Journal of Financial Intermediation 2, 168–206.

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

Goetzmann, W., Ingersoll, J., Spiegel, M., Welch, I., 2007. Portfolio performance manipulation and manipulation-proof performance measures. Review of Financial Studies 20, 1503–1546.

Goetzmann, W., Massa, M., 2002. Daily momentum and contrarian behavior of index fund investors. Journal of Financial and Quantitative Analysis 37, 375–389.

Goodhart, C., 1988. The foreign exchange market: A random walk with a dragging anchor. Economica 55, 437–460.

Griffin, J., Harris, J., Topaloglu, S., 2003. The dynamics of institutional and individual trading. Journal of Finance 58, 2285–2320.

Grinblatt, M., Keloharju, M., 2000. The investment behavior and performance of various investor types: A study of Finland's unique data set. Journal of Financial Economics 55, 43–67.

Grinblatt, M., Titman, S., Wermers, R., 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. American Economic Review 85, 1088–1105.

Hasbrouck, J., 1991. Measuring the information content of stock trades. Journal of Finance 46, 179–207.

Heston, S., Korajczyk, R., Sadka, R., 2010. Intraday patterns in the cross-section of stock returns. Journal of Finance 65, 1369–1407.

Hvidkjaer, S., 2006. A trade-based analysis of momentum. Review of Financial Studies 19, 457–491.

Kyle, A., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Lakonishok, J., Shleifer, A., Vishny, R., 1992. The impact of institutional trading on stock prices. Journal of Financial Economics 32, 23–43.

Latza, T., Payne, R., 2013. The information content of high-frequency limit order flows. Working Paper .

Linnainmaa, J., 2010. Do Limit Orders Alter Inferences about Investor Performance and Behavior? Journal of Finance 65, 1473–1506.

Linnainmaa, J.T., Saar, G., 2012. Lack of anonymity and the inference from order flow. Review of Financial Studies 25, 1414–1456.

Love, R., Payne, R., 2008. Macroeconomic news, order flows, and exchange rates. Journal of Financial and Quantitative Analysis 43, 467–488.

Lyons, R., 1997. A simultaneous trade model of the foreign exchange hot potato. Journal of International Economics 42, 275–298.

Marquering, W., Verbeek, M., 2004. The economic value of predicting stock index returns and volatility. Journal of Financial and Quantitative Analysis 39, 407–429.

Nolte, I., Nolte, S., 2010. How do individual investors trade? European Journal of Finance , forthcoming.

Nolte, I., Voev, V., 2011. Trading dynamics in the foreign exchange market: A latent factor panel intensity approach. Journal of Financial Econometrics 9, 685–716.

Pasquariello, P., Vega, C., 2007. Informed and strategic order flow in the bond markets. Review of Financial Studies 20, 1975–2019.

Payne, R., 2003. Information transmission in inter-dealer foreign exchange transactions. Journal of International Economics 61, 307–329.

Peiers, B., 1997. Informed traders, intervention, and price leadership: A deeper view of the microstructure of the foreign exchange market. Journal of Finance 52, 1589–1614.

Rime, D., Sarno, L., Sojli, E., 2010. Exchange rate forecasting, order flow and macroeconomic information. Journal of International Economics 80, 72–88.

Rindi, B., 2008. Informed traders as liquidity providers: Anonymity, liquidity and price formation. Review of Finance 12, 497.

Roell, A., 1990. Dual-capacity trading and the quality of the market. Journal of Financial Intermediation 1, 105–124.

Sapp, S., 2002. Price leadership in the spot foreign exchange market. Journal of Financial and Quantitative Analysis 37, 425–448.

Wermers, R., 1999. Mutual fund herding and the impact on stock prices. Journal of Finance 54, 581–622.

Appendix

Figure 1: Statistical Performance of the first (M1) Forecasting Model.

The graph gives an indication of the statistical performance of the first order flow model (M1, Equation 3) in the initial estimation window (03/30/2010–08/09/2010, 86 days). In particular, we report 2 summary statistics of the heteroskedasticity and autocorrelation corrected (Newey-West) t-statistics of the order flow (OF) coefficient. The first statistic is the number of positive coefficients (*#1* - blue bar) and the second statistic is the total number of significant coefficients (*#2* - red bar). OF is defined as the daily difference between the buyer-initiated and seller-initiated volume. Statistical significant is on a 10% level.

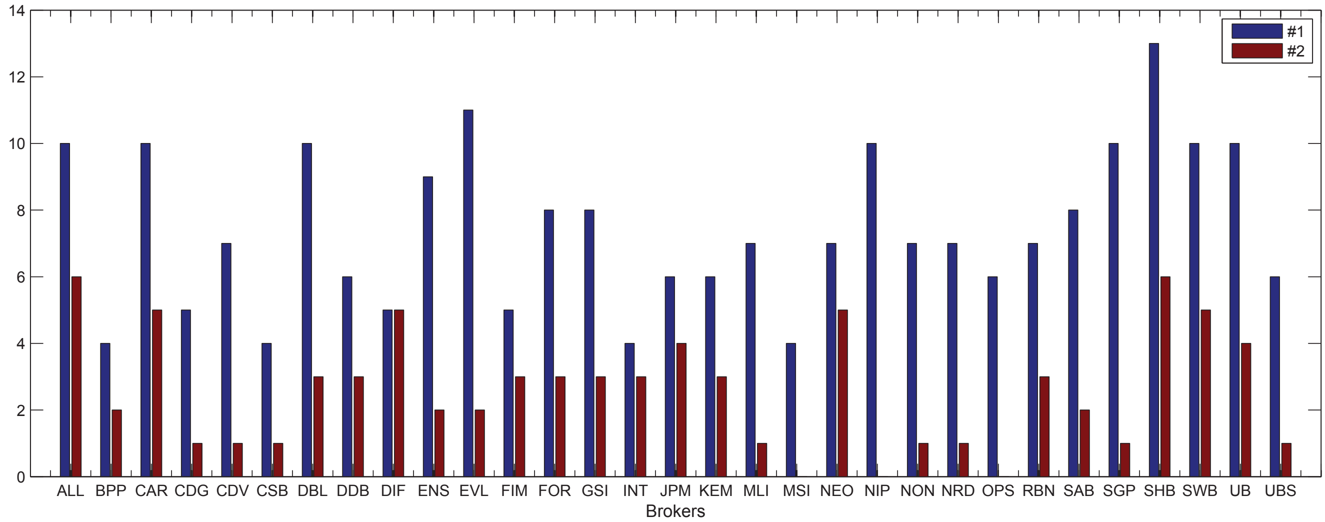
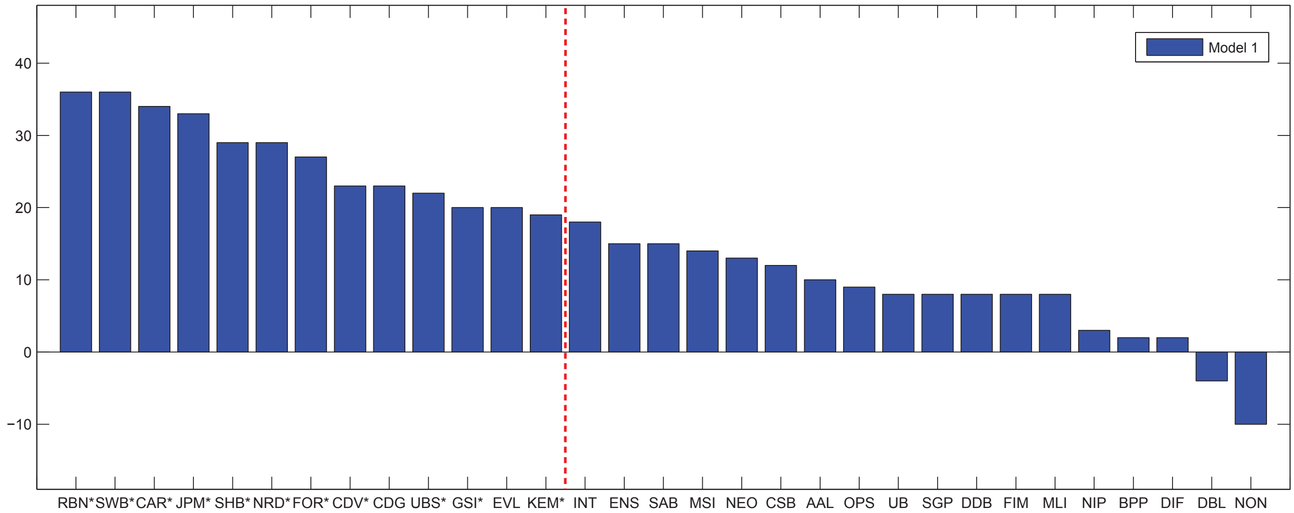
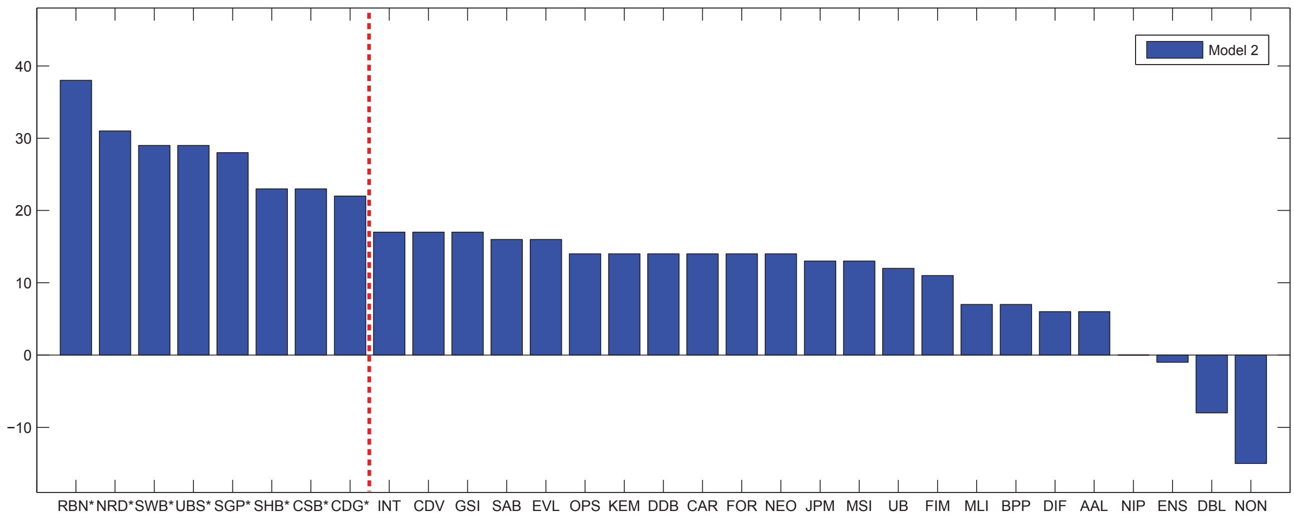


Figure 2: The Ranking of the Informativeness of Broker Identities.

The graph presents the out-of-sample performance of the mean-variance portfolios in descending order (from left to right) with respect to the Θ performance measure. We take the data from Table 5, where details on the methodology can be found. Briefly, the investment scenario is based on a risk-averse investor, who maximizes his expected portfolio return subject to an annual target volatility $\sigma_p = 10\%$. Every day, the investor forecasts next day's returns using the order flow models (M1 and M2) described in Section 3.1, and then rebalances his portfolio weights. There is one portfolio for each broker (x-axis). In Panel A (Panel B) we present the ranking of brokers' portfolios for M1 (M2). * indicates that the associated Θ is statistical significant, p-value is on a 10% level. The red line separates the group of significant and insignificant brokers.



(a)



(b)

Table 1: List of Brokers in HEX

These are the brokers of HEX ranked alphabetically. The first 2 columns are the full name and the code of brokers, respectively. The third column is the nationality of brokers. The letter R in the last column identifies the remote members. (Source: http://nordic.nasdaqomxtrader.com/membershipservices/membershiplist)

Broker Name	*Code*	*Country*	*RM*
Ålandsbanken Abp	AAL	Finland	
ABN AMRO Clearing Bank N.V	FOR	Netherlands	R
BNP Paribas Arbitrage SNC	BPP	France	R
Carnegie Investment Bank AB	CAR	Sweden	
Citadel Securities (Europe) Limited	CDG	UK	R
Citigroup Global Markets Limited	SAB	UK	R
Crédit Agricole Cheuvreux Nordic AB	CDV	Sweden	
Credit Suisse Securities (Europe) Ltd	CSB	UK	R
Danske Bank A/S	DDB	Denmark	
Deutsche Bank AG	DBL	UK	R
Evli Bank Abp	EVL	Finland	
FIM Bank Ltd.	FIM	Finland	
Goldman Sachs International	GSI	UK	R
Instinet Europe Limited	INT	UK	R
JP Morgan Securities Ltd	JPM	UK	R
Knight Capital Europe Limited	KEM	UK	R
Merrill Lynch International	MLI	UK	R
Morgan Stanley Co. International Ltd.	MSI	UK	R
NeoNet Securities AB	NEO	Sweden	
Nomura International plc	NIP	UK	R
Nordea Bank Finland Plc	NRD	Finland	
Nordnet Bank AB	NON	Sweden	
Pohjola Bank Plc	OPS	Finland	
SAXO-E*TRADE Bank A/S	DIF	Denmark	
Skandinaviska Enskilda Banken AB	ENS	Sweden	
Société Générale S.A.	SGP	France	R
Swedbank AB	SWB	Sweden	
Svenska Handelsbanken AB	SHB	Sweden	
The Royal Bank of Scotland N.V.	RBN	UK	R
UBS Limited	UBS	UK	R
UB Securities Limited	UB	Finland	

Table 2: Summary Statistics.

The table reports the name, mean (%), standard deviation (%), maximum (%), minimum (%), skewness, and kurtosis of the returns of the 15 most liquid stocks of HEX. All statistics are calculated over the daily interval for the whole sample period 03/30/2010 - 02/28/2011. The last column, reports the total aggressive turnover ('000,000) traded in each stock during the whole sample period.

		03/30/2010 - 02/28/2011						
#	*Name*	*Mean*	*Std*	*Max*	*Min*	*Skew*	*Kurt*	*Turn*
1	***Elisa***	0.03	1.36	7.08	-6.23	0.40	8.19	1,624
2	***Fortum***	0.09	1.34	4.75	-5.60	-0.28	4.81	7,059
3	***Kone***	0.11	1.57	5.77	-4.35	0.20	4.08	3,489
4	***Konecranes***	0.16	2.05	8.74	-4.69	0.78	5.23	1,641
5	***Metso***	0.18	2.49	9.35	-6.68	0.14	3.76	5,090
6	***Neste Oil***	-0.01	1.71	6.77	-6.65	-0.23	4.66	2,318
7	***Nokia***	-0.26	2.36	6.09	-15.33	-2.45	16.71	38,370
8	***Nokian Renkaat***	0.18	2.06	9.88	-5.89	0.58	5.07	2,990
9	***Outokumpu***	-0.10	2.15	8.44	-6.63	0.23	4.26	3,374
10	***Outotec***	0.18	2.51	11.40	-8.14	0.28	4.59	2,616
11	***Rautarukki K***	0.01	2.15	8.60	-5.85	0.49	4.39	1,978
12	***Sampo A***	0.06	1.53	8.89	-6.01	0.51	8.12	5,267
13	***Stora Enso R***	0.16	2.23	8.07	-6.68	0.12	3.86	6,095
14	***UPM-Kymmene***	0.16	2.04	8.44	-5.76	0.07	4.27	6,512
15	***Wärtsilä Abp***	0.17	2.08	9.61	-5.24	0.32	4.78	3,280

Table 3: A 5 Second Slice of the Transaction Data of Nokia

This is a 5 second slice of the transaction data of Nokia. The first two columns are the date and time expressed as month/day/year and hour:minute:second, respectively. The third column is the type of transaction, which can be Best Bid, Best Ask or Trade. The next two columns are the price (in euros) and the size of transaction. The last two columns are the Broker Buy Code and the Broker Sell Code.

Date	*Time*	*Type*	*Price*	*Size*	*Broker Buy*	*Broker Sell*
4012010	08:03:51	BEST BID	11.6	15,531		
4012010	08:03:51	BEST BID	11.6	13,531		
4012010	08:03:53	BEST ASK	11.61	14,876		
4012010	08:03:55	BEST ASK	11.61	9,876		
4012010	08:03:55	BEST ASK	11.61	6,876		
4012010	08:03:55	BEST BID	11.6	12,331		
4012010	08:03:55	TRADE	11.61	1,161	ENS	ENS
4012010	08:03:55	TRADE	11.61	39	ENS	NON
4012010	08:03:55	BEST ASK	11.61	5,676		
4012010	08:03:56	BEST BID	11.6	12,305		
4012010	08:03:56	TRADE	11.6	26	ENS	NON
4012010	08:03:56	TRADE	11.6	1,305	ENS	NON
4012010	08:03:56	TRADE	11.6	669	ENS	NON

Table 4: Performance in the Initial Period.

The table presents the performance of the mean variance portfolios in the initial period: 03/30/2010–08/09/2010 (86 days). There is one portfolio for each broker. The ANON portfolio is the one that disregards the broker identity, and it is the benchmark portfolio. The investment scenario is based on a risk-averse investor, who maximizes his expected portfolio return subject to an annual target volatility $\sigma_p = 10\%$. Every day, the investor forecasts next day's returns using the order flow models $M1 - M2$ described in Subsection 3.1, and then rebalances his portfolio weights. The order flow models are estimated once using all 86 days. We present: the annualized Sharpe ratio (SR) of each portfolio and Θ. Θ is the difference between the broker^j's and the ANON's performance measure (MPPM) of Goetzmann et al. (2007), which is expressed in (annualized) percentage points and is for $\gamma = 6$. It can be viewed as the maximum performance fee an investor is willing to pay to switch from the ANON portfolio to the broker^j's portfolio. When $\Theta > 0$, market transparency yields positive economic value to mean variance investors. Following Goetzmann et al. (2007), we test whether the broker^j's portfolio significantly outperforms the ANON portfolio and report the p-values in square brackets. We bold significant Θ on a 10% level.

	M1			M2		
Broker	*SR*	Θ	*p-val*	*SR*	Θ	*p-val*
AAL	4.82	-1	[0.53]	5.57	2	[0.46]
BPP	5.24	11	[0.32]	6.07	12	[0.29]
CAR	4.66	-8	[0.66]	5.08	-5	[0.61]
CDG	6.23	13	[0.26]	8.39	**32**	[0.07]
CDV	3.46	-16	[0.83]	5.55	2	[0.45]
CSB	4.75	1	[0.48]	4.58	2	[0.46]
DBL	7.98	**28**	[0.06]	8.43	**31**	[0.05]
DDB	7.43	**34**	[0.06]	7.67	**40**	[0.04]
DIF	5.44	-8	[0.67]	6.47	3	[0.43]
ENS	5.66	5	[0.40]	6.51	19	[0.23]
EVL	4.04	-22	[0.91]	4.35	-16	[0.80]
FIM	4.42	-7	[0.64]	4.87	-7	[0.65]
FOR	7.62	24	[0.12]	8.61	20	[0.12]
GSI	3.89	-17	[0.83]	5.62	-1	[0.53]
INT	5.99	8	[0.35]	5.48	0	[0.50]
JPM	5.25	11	[0.27]	7.22	**29**	[0.09]
KEM	5.30	-1	[0.52]	5.85	-1	[0.52]
MLI	5.65	-1	[0.51]	5.17	-6	[0.62]
MSI	6.82	18	[0.15]	6.13	5	[0.40]
NEO	5.33	-4	[0.58]	6.16	5	[0.40]
NIP	4.96	-2	[0.53]	6.25	20	[0.19]
NON	5.67	14	[0.26]	6.94	29	[0.11]
NRD	6.86	23	[0.11]	8.32	**30**	[0.06]
OPS	3.06	-20	[0.79]	4.05	-13	[0.72]
RBN	5.59	8	[0.36]	6.25	16	[0.23]
SAB	5.94	8	[0.35]	6.79	13	[0.26]
SGP	4.46	-11	[0.73]	4.95	-6	[0.63]
SHB	6.91	22	[0.16]	5.82	10	[0.34]
SWB	5.78	15	[0.21]	7.84	**41**	[0.03]
UB	5.87	11	[0.30]	6.59	14	[0.26]
UBS	5.55	3	[0.44]	7.41	16	[0.17]
ANON	5.42	0	[0.50]	5.42	0	[0.50]

Table 5: Does Broker Identity Convey Information?

The table presents the performance of the mean-variance portfolios, using a recursive (out-of-sample) regression estimation, which is based on a window of expanding size. The period is 08/10/2010–02/28/2011 (124 days). There is one portfolio for each broker. The ANON portfolio is the one that disregards the broker identity, and it is the benchmark portfolio. The investment scenario is based on a risk-averse investor, who maximizes his expected portfolio return subject to an annual target volatility $\sigma_p = 10\%$. Every day, the investor forecasts next day's returns using the order flow models $M1 - M2$ described in Subsection 3.1, and then rebalances his portfolio weights. We present: the annualized Sharpe ratio (SR) of each portfolio and Θ. Θ is the difference between the broker^j's and the ANON's performance measure (MPPM) of Goetzmann et al. (2007), which is expressed in (annualized) percentage points and is for $\gamma = 6$. It can be viewed as the maximum performance fee an investor is willing to pay to switch from the ANON portfolio to the broker^j's portfolio. When $\Theta > 0$, market transparency yields positive economic value to mean variance investors. Following Goetzmann et al. (2007), we test whether the broker^j's portfolio significantly outperforms the one that disregards the broker identity and report the p-values in square brackets. We bold significant Θ on a 10% level.

	M1			M2		
Broker	*SR*	Θ	*p-val*	*SR*	Θ	*p-val*
AAL	0.37	10	[0.29]	-0.05	6	[0.38]
BPP	-0.47	2	[0.44]	0.00	7	[0.33]
CAR	2.93	**34**	[0.01]	0.72	14	[0.20]
CDG	1.41	23	[0.14]	1.66	22	[0.11]
CDV	1.45	**23**	[0.08]	0.94	17	[0.17]
CSB	0.51	12	[0.23]	1.45	**23**	[0.09]
DBL	-0.85	-4	[0.59]	-1.11	-8	[0.67]
DDB	0.18	8	[0.31]	0.72	14	[0.20]
DIF	-0.55	2	[0.45]	-0.09	6	[0.36]
ENS	0.81	15	[0.12]	-0.79	-1	[0.53]
EVL	1.29	20	[0.15]	0.88	16	[0.19]
FIM	0.17	8	[0.34]	0.47	11	[0.27]
FOR	2.00	**27**	[0.04]	0.71	14	[0.17]
GSI	1.33	**20**	[0.09]	1.01	17	[0.15]
INT	1.03	18	[0.13]	0.97	17	[0.14]
JPM	2.21	**33**	[0.04]	0.65	13	[0.24]
KEM	1.23	**19**	[0.08]	0.75	14	[0.18]
MLI	0.10	8	[0.32]	0.01	7	[0.33]
MSI	0.75	14	[0.21]	0.60	13	[0.27]
NEO	0.63	13	[0.17]	0.70	14	[0.20]
NIP	-0.22	3	[0.42]	-0.49	0	[0.49]
NON	-1.61	-10	[0.73]	-2.04	-15	[0.80]
NRD	2.44	**29**	[0.02]	2.79	**31**	[0.03]
OPS	0.30	9	[0.29]	0.74	14	[0.22]
RBN	2.64	**36**	[0.03]	2.90	**38**	[0.03]
SAB	0.81	15	[0.22]	0.90	16	[0.19]
SGP	0.19	8	[0.31]	2.14	**28**	[0.07]
SHB	2.23	**29**	[0.03]	1.61	**23**	[0.09]
SWB	2.50	**36**	[0.03]	1.93	**29**	[0.06]
UB	0.16	8	[0.29]	0.54	12	[0.23]
UBS	1.42	**22**	[0.10]	2.08	**29**	[0.08]
ANON	-0.57	0	[0.50]	-0.57	0	[0.50]

Table 6: Brokers' Market Share Statistics.

The table presents the market share statistics of brokers in HEX in the period 03/29/2010-02/28/2011. The first statistic (*Vol(1) %*) is the market share with respect to the average daily volume initiated by each broker across the 15 most liquid stocks of HEX. The second statistic (*Vol(2) %*) is the market share with respect to the average daily volume executed by each broker. We boldface the larger brokers (Q4) for each market share statistic.

Broker	*Vol(1) %*	*Vol(2) %*
AAL	*0.5%*	*0.6%*
BPP	*4.1%*	*4.2%*
CAR	*1.8%*	*2.2%*
CDG	***10.5%***	***5.8%***
CDV	*1.7%*	*1.8%*
CSB	***7.9%***	***7.0%***
DBL	*3.6%*	***4.7%***
DDB	***4.3%***	*3.7%*
DIF	*0.5%*	*0.3%*
ENS	***8.3%***	***9.0%***
EVL	*1.1%*	*1.2%*
FIM	*4.2%*	*4.2%*
FOR	***6.3%***	***7.1%***
GSI	*2.5%*	*2.7%*
INT	*0.8%*	*0.8%*
JPM	*1.6%*	*1.8%*
KEM	*0.4%*	*0.3%*
MLI	*2.8%*	*3.1%*
MSI	*3.2%*	***4.7%***
NEO	*1.0%*	*1.0%*
NIP	***4.4%***	*3.0%*
NON	*4.0%*	*4.4%*
NRD	***5.2%***	***5.6%***
OPS	*2.3%*	*3.0%*
RBN	*0.9%*	*1.0%*
SAB	*2.9%*	*3.1%*
SGP	***6.4%***	***4.6%***
SHB	*2.5%*	*3.4%*
SWB	*1.7%*	*2.0%*
UB	*0.6%*	*1.1%*
UBS	*2.0%*	*2.7%*

Table 7: Do Large Brokers Outperform Small?

The table presents the performance difference between a portfolio that tracks large brokers and one that tracks small brokers. We use 2 market share criteria to categorize brokers: a. with respect to the average daily volume initiated, *Vol(1)*, and b. the average daily volume executed, *Vol(2)*. Next, we construct daily average order flow measure series of the top (large brokers - Q4) and bottom (small brokers - Q1) quartile of brokers for each criterion. We build daily rebalancing mean-variance portfolios using the order flow models $M1 - M2$, described in Subsection 3.1, to predict next day's returns. This out-of-sample recursive regression estimation is based on a window of expanding size in the period 08/10/2010–02/28/2011 (124 days). We estimate the MPPM of Goetzmann et al. (2007) for each group of brokers and report the performance of each portfolio against ANON and $\Delta\Theta$, which is the performance difference expressed in annual percentage points and for $\gamma = 6$. Following Goetzmann et al. (2007), we test whether the large brokers' portfolio significantly outperforms ($\Delta\Theta > 0$) the small brokers' portfolio and report the p-values in square brackets.

	M1	M2
a. Vol(1)		
Large Brokers	5	2
p-val	[0.36]	[0.45]
Small Brokers	39	44
p-val	[0.02]	[0.00]
$\boldsymbol{\Delta\Theta}$	-34	-42
p-val	[0.97]	[0.99]
b. Vol(2)		
Large Brokers	-13	-24
p-val	[0.79]	[0.93]
Small Brokers	38	44
p-val	[0.02]	[0.00]
$\boldsymbol{\Delta\Theta}$	-51	-68
p-val	[0.99]	[1.00]

Table 8: Brokers' Investor Base Heterogeneity, Market Share, and Portfolio Performance.

In panel a., the table presents the correlated trading statistics of the daily trades of brokers in HEX in the period 03/29/2010-28/02/2011. We measure correlated trading by the herding measure (LSV) of Lakonishok et al. (1992), which is defined in Equation 9. The LSV statistics are computed for each stock-day and then averaged. If trades are independent, the mean LSV measure will be zero. In panel b., we present the average size statistics of the quartile of brokers with the highest (*High*) and lowest (*Low*) LSV statistic. The size is measured with respect to the average daily volume initiated (*Vol(1)*) or executed (*Vol(2)*, aggressive and passive) by each broker. In panel c., we construct daily average order flow measure series of the top (Q4) and bottom (Q1) quartile of brokers, and then we build daily-rebalancing mean-variance portfolios, using the order flow models $M1 - M2$, described in Subsection 3.1, to predict next day's returns. This out-of-sample recursive regression estimation is based on a window of expanding size in the period 08/10/2010–02/28/2011 (124 days). We estimate the performance measure of Goetzmann et al. (2007) of each quartile of brokers and report $\Delta\hat{\Theta}$, which is the performance difference between the brokers with high (Q4) and low (Q1) correlated trading, expressed in annual percentage points and for $\gamma = 6$. Following Goetzmann et al. (2007), we test whether the top quartile of brokers significantly outperforms ($\Delta\Theta > 0$) the bottom quartile. We report the standard deviation (*std*) of $\Delta\Theta$, and the p-values in square brackets.

a. LSV Statistics

Brokers	*LSV*	*Brokers*	*LSV*	*Brokers*	*LSV*
AAL	0.10	***FIM***	0.05	***NRD***	0.11
BPP	0.17	***FOR***	0.05	***OPS***	0.10
CAR	0.18	***GSI***	0.15	***RBN***	0.27
CDG	0.02	***INT***	0.23	***SAB***	0.14
CDV	0.23	***JPM***	0.24	***SGP***	0.08
CSB	0.09	***KEM***	0.23	***SHB***	0.16
DBL	0.13	***MLI***	0.11	***SWB***	0.14
DDB	0.09	***MSI***	0.19	***UB***	0.14
DIF	0.03	***NEO***	0.12	***UBS***	0.18
ENS	0.15	***NIP***	0.15		
EVL	0.20	***NON***	0.05		

b. LSV and Brokers' Size

LSV	*Vol(1)*	*Vol(2)*
High (Q4)	32,073	79,874
Low (Q1)	120,689	206,390

c. LSV and Mean-Variance Performance

	M1	M2
$\boldsymbol{\Delta\Theta}$	26	25
p-val	[0.08]	[0.13]

Table 9: Active Brokers as Proxy for Sophistication

The table presents the performance of a mean-variance portfolio that uses the average order flow measure of the most active brokers (Q4 quartile) at time t to predict returns at time $t+1$ (Panel a.), $t+2$ (Panel b.), and $t+3$ (Panel c.) using the order flow models $M1-M2$, described in Subsection 3.1. We rebalance portfolio's weights on a daily frequency. This out-of-sample recursive regression estimation is based on a window of expanding size in the period 08/10/2010–02/28/2011 (124 days). We estimate the MPPM of Goetzmann et al. (2007) and report the performance difference, Θ, against the portfolio that disregards the broker identity (ANON). Θ is expressed in (annualized) percentage points and is for $\gamma = 6$. When $\Theta > 0$, market transparency yields positive economic value to mean-variance investors. Following Goetzmann et al. (2007), we test whether the portfolio significantly outperforms the ANON portfolio and report the p-values in square brackets.

	M1	M2
a. Use OF_t^i ***to predict*** R_{t+1}^i		
$\boldsymbol{\Theta}$	28	39
p-val	[0.00]	[0.00]
b. Use OF_t^i ***to predict*** R_{t+2}^i		
$\boldsymbol{\Theta}$	7	3
p-val	[0.30]	[0.39]
c. Use OF_t^i ***to predict*** R_{t+3}^i		
$\boldsymbol{\Theta}$	10	8
p-val	[0.29]	[0.30]

Table 10: Analysis of the Investment Style of Brokers.

The table presents the fraction of positive buy ratio differences across brokers (including the ANON portfolio) for the period 03/29/2010–02/28/2011. We follow Grinblatt and Keloharju (2000) to construct buy ratios: buy volume/(buy volume+sell volume). In our calculations we use only aggressive trades. Each buy ratio difference is generated by subtracting the average buy ratio of stocks in the bottom quartile (losers) from the average buy ratio of stocks in the top quartile (winners). We use hourly and daily buy ratios, while the past returns used for ranking the stocks are based on the previous hour and day, respectively. We present the fraction of positive buy ratio differences ($BRDif$). Under the hypothesis of no momentum or contrarian behavior, the average buy ratio difference should be zero, and the aforementioned fraction equal to 0.50. A fraction which is larger than 0.50 indicates a momentum trading behavior, while a fraction smaller than 0.50 indicates a contrarian behavior. In square brackets we report the p-values of the standard binomial test ($p-val$) of whether the fraction of buy ratio differences is 0.50. We drop zero buy ratio differences from the fraction calculation. We bold significant $BRDif$ on a 10% level.

	1 hour		1 day	
Broker	*BRDif*	*p-val*	*BRDif*	*p-val*
AAL	0.49	[0.45]	**0.39**	[0.00]
BPP	0.49	[0.59]	0.50	[0.90]
CAR	0.49	[0.59]	0.50	[0.95]
CDG	**0.60**	[0.00]	0.55	[0.13]
CDV	**0.57**	[0.00]	0.50	[0.89]
CSB	**0.43**	[0.00]	**0.57**	[0.04]
DBL	**0.42**	[0.00]	**0.28**	[0.00]
DDB	**0.42**	[0.00]	0.52	[0.47]
DIF	**0.46**	[0.00]	**0.57**	[0.03]
ENS	**0.43**	[0.00]	**0.41**	[0.01]
EVL	0.49	[0.30]	0.45	[0.15]
FIM	**0.40**	[0.00]	0.49	[0.74]
FOR	**0.58**	[0.00]	**0.65**	[0.00]
GSI	**0.55**	[0.00]	**0.59**	[0.01]
INT	**0.56**	[0.00]	**0.58**	[0.02]
JPM	**0.59**	[0.00]	0.55	[0.17]
KEM	**0.60**	[0.00]	0.56	[0.13]
MLI	0.52	[0.13]	0.49	[0.74]
MSI	**0.43**	[0.00]	**0.41**	[0.00]
NEO	**0.53**	[0.02]	**0.57**	[0.04]
NIP	**0.55**	[0.00]	0.51	[0.74]
NON	**0.34**	[0.00]	0.53	[0.39]
NRD	**0.40**	[0.00]	**0.35**	[0.00]
OPS	**0.41**	[0.00]	**0.39**	[0.00]
RBN	**0.55**	[0.00]	0.53	[0.33]
SAB	**0.60**	[0.00]	**0.68**	[0.00]
SGP	**0.48**	[0.04]	0.48	[0.47]
SHB	0.51	[0.38]	0.55	[0.17]
SWB	**0.53**	[0.04]	**0.58**	[0.02]
UB	0.47	[0.06]	0.53	[0.44]
UBS	**0.58**	[0.00]	**0.59**	[0.00]

Table 11: Do Momentum Brokers Outperform Contrarian?

The table presents the performance of a mean-variance portfolio that uses the average order flow of momentum brokers at time t to predict returns at time t+1 using the order flow models M1 - M2, described in Subsection 3.1. We repeat for contrarian brokers. We follow Grinblatt and Keloharju (2000) to characterize brokers as momentum or contrarian. Momentum are the brokers with daily buy ratio difference fraction significantly (p-value$<$ 5%) greater than 0.50. Contrarian are the brokers with daily buy ratio difference fraction significantly (p-value$<$ 5%) smaller than 0.50. We rebalance portfolio's weights on a daily frequency. This out-of-sample recursive regression estimation is based on a window of expanding size in the period 08/10/2010–02/28/2011 (124 days). We estimate the MPPM of Goetzmann et al. (2007) and report the performance difference of the two portfolios against the one that disregards the broker identity (ANON). We also report $\Delta\Theta$, which is the performance difference between the momentum and contrarian portfolio. Performance differences are expressed in annual percentage points and are for $\gamma = 6$. Following Goetzmann et al. (2007), we test if the two portfolios significantly outperform ($\Theta > 0$) the ANON portfolio, as well as if the momentum portfolio significantly outperforms ($\Delta\Theta > 0$) the contrarian. We report the p-values in square brackets.

	M1	M2
Momentum	15	30
p-val	[0.11]	[0.01]
Contrarian	6	2
p-val	[0.37]	[0.46]
$\boldsymbol{\Delta\Theta}$	9	28
p-val	[0.29]	[0.04]

Table 12: Stock Picking Ability and Investment Style

The table shows the relation between stock picking ability and investment style in the period 03/29/2010–02/28/2011. We follow Grinblatt and Keloharju (2000) to construct buy ratio difference fractions based on future one- and three-months returns in order to measure brokers' stock picking ability. In the absence of stock picking ability, the average buy ratio difference should be zero, and the aforementioned fraction equal to 0.50. A fraction larger (smaller) than 0.50 means that the stocks brokers buy on a daily basis have a positive (negative) one- or three-months performance, thus, brokers have high (low) stock picking ability. We split brokers into two groups; those with high stock picking ability (Q4 quartile), and those with low stock picking ability (Q1 quartile). We, then, follow Grinblatt and Keloharju (2000) to measure brokers' investment style and report the relevant average buy ratio difference fraction ($BRDif$) of each group based on one-day past returns. A fraction which is larger than 0.50 indicates a momentum trading behavior, while a fraction smaller than 0.50 indicates a contrarian behavior. We, also, report the difference of investment styles of the two groups and the associated p-value in square brackets.

Stock Picking Ability	*BRDif*	*Investment Style*
a. 1 month		
High (Q4)	0.57	*Momentum*
Low (Q1)	0.48	*Contrarian*
Q4-Q1	0.09	
p-val	[0.00]	
b. 3 months		
High (Q4)	0.56	*Momentum*
Low (Q1)	0.46	*Contrarian*
Q4-Q1	0.10	
p-val	[0.00]	