Whose trades convey information?

Lukas Menkhoff, University of Hannover, Germany
and
Maik Schmeling, University of Hannover, Germany

Abstract:
Information is a property of *traders* and not necessarily of *trades*. Accordingly, we analyze *traders*' characteristics in an electronic limit order market via anonymous trader identities. We use six indicators of informed trading in a cross-sectional multivariate approach to identify traders with high price impact. More information is conveyed by those traders' trades who – simultaneously – use medium-sized orders (practice stealth trading), have large trading volume, are located in a financial center, trade early in the trading session, at times of wide spreads and when the order book is thin. These variables tentatively have a declining marginal effect on identification of informed traders.

JEL-Classification: G12, G15, D82, F31
Keywords: Market microstructure, informed trading, local proximity, trade size, foreign exchange

April 28, 2006

We would like to thank the German Research Foundation (Deutsche Forschungsgemeinschaft DFG) for financial support. We gratefully acknowledge research assistance by Leila Gadijeva.

corresponding: Lukas Menkhoff, Department of Economics, University of Hannover, Königsworther Platz 1, D-30167 Hannover, Germany, menkhoff@gif.uni-hannover.de

Whose trades convey information?

1 Introduction

There is strong evidence that information in financial markets is aggregated via the trading process. Information in market transactions can be identified due to its permanent price impact, whereas other effects on prices, such as from balancing inventory holdings, will have transient price impact only (Hasbrouck, 1991, 1991a, 2006). Consequently, every trade will have inventory effects but not necessarily information effects.[1] This naturally raises the question: whose trades convey information? This question is at the heart of the information aggregation process because it must be market participants who have information and trade on this information.

Despite its core importance and seeming simplicity, the identification of informed traders is still listed among the big open questions in the microstructure literature (Lyons, 2001, Hasbrouck, 2006). Obviously, the reason why informed traders are so difficult to identify is the limitation of available data. In an optimum setting one would be able to trace each traders' trades in the whole market (segment) including every single characteristic of these trades.[2] In reality, however, studies can only approximate this optimum – the main bottleneck in empirical work is trader identity. Accordingly, most studies in this field circumvent the focus on *traders* and instead analyze *trades*. They observe the time-series dimension and find (bivariate) relations of higher price impacts with some trade characteristics. These characteristics are interpreted from the viewpoint of information processing, such that larger trade size or wider bid-ask spread convey information (Hasbrouck, 1991, Koski and Michaely, 2000, Hasbrouck and Seppi, 2001, Payne, 2003, Brandt and Kavajecz, 2004, Bjønnes and Rime, 2005). However, this approach is necessarily second-best only because information is in the last instance not a property of a *trade* but of a *trader*. Our paper shows that this distinction is more than theoretical reasoning but makes a difference in the market we analyze.

The importance of this issue may be clarified by a simple example. Suppose there are two informed traders who want to capitalize on their private information. One of the traders

---

[1] For the information contained in order flow see for example Hasbrouck (1991, 1991a, 2006), Dufour and Engle (2000) or Dunne, Hau and Moore (2004) on stock markets, Brandt and Kavajecz (2004) and Green (2004) on bond markets and Ito, Lyons and Melvin (1998), Evans and Lyons (2002, 2002a, 2005, 2005a), Froot and Ramadorai (2005), Payne (2003) or Gehrig and Menkhoff (2004) on foreign exchange markets.

[2] There are further approaches to analyze asymmetric information, such as comparing financial versus non-financial customer order flow (Lyons, 2001, Bjønnes, Rime and Solheim, 2005) or examining profitability of position taking of market participants with and without local proximity to firms' headquarters (e.g. Coval and Moskowitz, 2001, Hau, 2001, Malloy, 2005).

uses few large orders whereas the other uses lots of small trades. Every attempt to empirically assess whether large or small trades are more informative will be severely hampered. Employing a trader-centered analysis, however, measures the average information content of all trades of a certain trader, i.e. his small, medium and large trades alike.

To examine this issue we exploit anonymous trader identities on all trades in an electronic limit order interbank currency market. This unique data enables us to characterize those traders whose trades have a high permanent price impact and in this way convey information. We conduct a cross-sectional regression of traders' price impact on six traders' trade characteristics as right-hand side variables. Thus, we complement the conventional analysis of *trade* type, such as large vs. small trades, by a *trader*-centered analysis, such as traders using large or small trades. Moreover, the consideration of six information indicators in a single regression allows inferences about relative importance of these indicators. We control, for example, for the fact whether effects from *trade* size may be influenced by the overall size of a *trader,* which is not necessarily the same.

It is our main finding that informed traders, i.e. traders whose market orders have a large permanent price impact, can be identified by the following six significant characteristics: information is conveyed by traders who – simultaneously – trade medium-sized orders, have large trading volume, are located in a financial center, trade early in the trading session, at times of wide spreads and when the order book is thin. It is interesting to see that this result is not necessarily the same when we use different methods. It is in particular the trade size variable whose indication of informed trade depends on the method applied, as we subsequently discuss.

Conventional theory states that larger trades should be more informative because traders who have information want to exploit this information quickly and thus issue larger orders. With the rise of limit order markets, however, this conventional wisdom has been questioned because trade sizes in limit order markets are very much standardized. If we thus run (i) a bivariate time-series correlation between price impact and order size for our limit order market we do not find a significant relation (consistent with Bjønnes and Rime, 2005). If we, however, run (ii) a bivariate cross-sectional correlation between traders' average order size and their price impacts we find a positive sign: informed traders tend to use larger orders. As a next step in our analysis, when we control trade size for other determinants (iii), its influence becomes insignificant again. Finally, if we analyze possible non-linear relations (iv) in a multivariate approach we find that informed traders use medium-sized orders, i.e. the result mentioned above.

Fortunately, results for the next four variables in the order of our discussion are robust to the method used: larger trader size, trader location in a financial center, earlier trading times and larger spreads are associated with higher permanent price impacts of a trader. It is the sixth indicator of information, the order book variable, which requires some discussion. While order book depth alleviates the price impact of a trade in a bivariate time-series setting, it is positive in our bivariate cross-sectional section. Thus, informed traders tend to trade when the book is thick. This is consistent with the idea of Admati and Pfleiderer (1988) that informed traders may try to hide their trades when there is lots of liquidity. However, in a multivariate regression – where we control for other indicators of information – the relation between traders' rice impacts and the average outstanding order book depth at which they trade turns negative. Obviously, the positive relation of informed trading with deep markets is driven by other factors, most probably by trader size and trader location.

In addition to identifying these determining characteristics of informed traders, the above analysis sheds light on several hypotheses about informed trade: (1) This paper adds to a few others showing that trader size matters for informed trade. (2) It shows that local information is important in foreign exchange even in a multivariate approach. (3) This study is among the first to test the Bloomfield, O'Hara and Saar (2005) hypothesis that trading time is important for the identification of informed traders with real market data.

(4) Beyond these tests of single hypotheses, our multivariate approach is informative about the relative economic significance of the six indicators of informed trade considered here. We find that the spread our dealers trade at has the largest effect on the information content of their trades as may be expected from the focus of the early microstructure literature on this important variable. The second most important variable is location, a finding also in line with many more recent studies identifying a local information advantage. The other variables lag behind these two, but still significantly contribute to our understanding of the informational role of order flow.

(5) In addition, we use quantile regressions to find out about possible non-linear effects in the characteristics of informed traders. We show a declining marginal impact of information indicators, such as that the first traders from a financial center increase informed trade more than the last. In a sense, we find that an information conveyance function is of Cobb-Douglas shape in the arguments trader size and trader location. (6) Finally, due to the importance of the trade size variable in the literature and our method-sensitive results, we use the highly disaggregated data to demonstrate the effect of so-called "stealth trading" in a more rigorous way than was possible before (Chakravarty, 2001).

This study is based on a new data base, i.e. an anonymous but otherwise complete record of transactions at a modern pure limit order market.[3] We cover nine days of Russian rouble–US dollar trading in 2002 at Moscow's MICEX exchange, the only countrywide platform of interbank trading. As this electronic market was newly designed in cooperation with established suppliers, it is no surprise that market characteristics closely mirror other limit order markets, such as the NYSE or US dollar–euro trading, despite the market's smaller size.

The rest of the paper is structured in five sections. Section 2 briefly reviews literature about indicators of informed trade from which we derive our six trader characteristics. An overview of the data and descriptive statistics employed is provided in Section 3. Results are presented in Section 4, robustness tests in Section 5 and conclusions in Section 6.

2 Literature and hypotheses

This section discusses six indicators of asymmetric information employed in earlier work. We are interested in their likely effect on the impact of order flow on prices.

The first variable of interest is (average) trade size of a trader, which is commonly taken to be an important indicator of informed trade. In traditional microstructure, a larger trade size is typically seen as carrying more information since informed traders will try to trade larger quantities to capitalize on their private information (see e.g. Kyle, 1985, Easley and O'Hara, 1987, Madhavan and Smidt, 1991). Therefore, one would expect a positive relation between mean trade size of a trader and his order flow's price impact. Bjønnes and Rime (2005) indeed find larger trades to be more informative in a setting of direct bilateral trades. However, this traditional indicator of informed trade becomes questionable in modern limit order markets which make it easy for informed traders to split their orders, thereby hiding their intended trade size (see Bernhardt and Hughson, 1997, or Chordia and Subrahmanyam, 2004, for a discussion of "splitting orders"). Chakravarty (2001) empirically investigates the effect of "stealth trading" (Barclay and Warner, 1993) and finds that medium sized trades have the highest price impact. The relationship between mean trade size of a market participant and his order flow's price impact in a limit order book therefore is a priori ambiguous.

---

[3] Electronic limit order book have gained a lot of attention in the empirical and theoretical literature since they are becoming the dominant trading environment for most kinds of assets (see inter alia Glosten, 1994, Biais, Hillion and Spatt, 1995, Evans, 2002 or Hollifield, Miller and Sandås, 2004).

A second and closely related variable is trader size.[4] There is little direct evidence on this issue for equity markets where the discussion is rather framed in terms of (small) individual vs. (large) institutional traders (see Campbell, Ramadorai and Vuolteenaho, 2005, for a discussion of these procedures). Evidence shows that large traders (institutional investors) possess superior information compared to small traders (individuals) (Chakravarty, 2001, Jones and Lipson, 2004, Sias and Starks, 1997, Sias, Starks and Titman, 2004). In foreign exchange markets, large traders are typically viewed as possessing superior information since they have a larger customer base, which is the main source of private information for foreign exchange dealers (see Evans and Lyons, 2005a). Furthermore, market participants do actually believe that large players are more informed (see Cheung and Chinn, 2001). Early evidence on a potentially important role of larger traders was provided by Peiers (1997) in an analysis of the role of single large banks in leading the market. Due to these results we expect order flow from large traders to have higher price impact.

Finally, a third potentially important variable is local proximity of a trader to a financial or economic center. Empirical evidence (see inter alia Coval and Moskowitz, 2001, Hau, 2001, Malloy, 2005) forcefully indicates that local proximity to corporate headquarters provides an informational edge for mutual fund managers, traders, or analysts. For foreign exchange markets, Covrig and Melvin (2002) show that Japanese traders tend to lead the Yen market. With a focus on end-users as the true source of information, financial customers are regarded as better informed (Lyons, 2001, Marsh and O'Rourke, 2005, Osler, Mende and Menkhoff, 2006). We therefore expect traders from financial centers of a country to have superior information since they have better access to order flow from financial customers. Hence, price impact and local proximity should be positively correlated.

The time of day at which a trader places his orders is a fourth potentially important determinant of price impacts. Bloomfield, O'Hara and Saar (2005) experimentally show that informed traders tend to trade early in a trading session to capitalize on their private information. Therefore, earlier trading is a proxy for superior information and we expect price impacts to be the higher the earlier a trader places his orders (on average).

A key variable in microstructure is the bid ask spread, which was originally considered as a measure of transaction costs and compensation for holding inventories (see inter alia Demsetz, 1968, and Ho and Stoll, 1981). Subsequent work also pointed out the importance of

[4] This is not to be confused with (mean) trade size of a trader as discussed in the paragraph above. A trader who transacts large quantities in total may do this e.g. with a sequence of small trades or with one large trade. Therefore, average trade size of a trader and his overall size are not necessarily the same.

spreads to cover costs associated with adverse selection (see e.g. Copeland and Galai, 1983, Glosten and Milgrom, 1985, Easley and O'Hara, 1987) when market makers are exposed to informed trade. Huang and Stoll (1997) empirically decompose spreads in equity markets and show that bid ask spreads indeed cover order processing costs, inventory costs and rewards for adverse selection. Payne (2003) for order driven and Osler, Mende and Menkhoff (2006) for quote driven markets also find spreads to compensate for adverse selection in currency markets. Based on these earlier findings, we expect the spread at which traders place their orders to positively influence the overall price impact of orders.

The last variable of interest is the outstanding order book volume. Higher liquidity naturally alleviates the total price impact of order flow. Here, however, we analyze exclusively the permanent price impact so that liquidity effects should be neglected. In the model of Admati and Pfleiderer (1988) liquidity traders decide to trade together to guard against the informed. Therefore, one should expect to find a negative relation between the level of liquidity a trader prefers to trade at and his price impact (see Payne, 2003). However, as also discussed in Admati and Pfleiderer (1988), higher liquidity could attract informed traders attempting to trade at low costs. Traders who trade when order book volume is comparably high might therefore show up with higher price impacts in a cross-sectional analysis. Therefore, the relation between the level of liquidity at which a trader prefers to trade and his price impact seems a priori unclear.

Each of these six indicators seems to be important to understand whose trades convey information in the market. It is the advantage of our data that this full set of indicators can be examined in a single framework.

3 Market structure, data, and descriptive statistics

3.1 Data set and market structure

Our data set covers spot RUR/USD trading at the MICEX in Moscow from March 11 to March 21, 2002 which took place in the so called Unified Trading Session (UTS).[5] The UTS was initiated by the Russian central bank and serves as a country-wide clearing for traders from all over Russia. This is important because in 2002 electronic currency trading in Russia took place on eight regional exchanges which were not linked to each other. Therefore, traders from e.g. Moscow could not trade electronically with traders from e.g. Rostov. To overcome this regional detachment, the Russian central bank encouraged a centralized interbank

[5] Goldberg and Tenorio (1994) also analyze trading at the MICEX. However, their data do not come from the modern electronic trading system.

market, the UTS, which accounts for the largest share of Russia's electronic currency trading. The importance of trading at this unified session also stems from the fact that the resulting price from the UTS serves to fix Russia's official exchange rate. Accordingly, the eight local currency exchanges follow the UTS rate closely.

Trading takes place in an electronic limit order book called SELT which is very similar to the systems of Reuters and EBS.[6] At the time in 2002, trading at the UTS took place one hour per day, starting 10.30 Moscow time. Nowadays, trading also takes place electronically in RUR/EUR and is prolonged to three hours per day.

Traders can submit limit orders and cancel any of their outstanding orders continuously during the trading session. Market orders can easily be constructed by submitting marketable limit orders.[7] The limit order book has clear time and price priority rules as encountered on virtually all modern trading systems. Marketable limit orders are executed immediately. Non-marketable limit orders are stored in the book and arranged by their associated price. If there is more than one limit order for a given price, the earlier submitted order has priority when hit by a crossing limit order.

The trading screen shows the best bid and ask prices with corresponding volumes. Furthermore, information about the last trade (size and direction) and cumulative trade size for buy and sell orders of the actual session are displayed. Finally, trading is anonymous and trader identification is revealed only to counterparties after completing a trade.

We have data on all activities in the limit order book and thus do not need to apply any kind of trade classification algorithm. These data also allow a precise reconstruction of the state of the limit order book in event time. Furthermore, we have coded trader identities for the whole population of traders which enable us to follow single traders through all their activities in SELT.

Table 1 shows some descriptive statistics for our trading data from the UTS. The data set contains 14,109 transactions with a mean transaction size of about 50,000 USD, which is much lower than e.g. in the EUR/USD market, which reports mean trade size of more than one million USD. The percentage spread averages 0.0071, which is also somewhat smaller than in the EUR/USD market (see Payne, 2003).

Looking at market statistics for the twelve non-overlapping five minute intervals that make up the one hour UTS trading, we find that spreads follow the well-known U-shaped pattern and we also observe the familiar inverted U-shaped pattern for outstanding limit or-

[6] SELT was in fact developed in cooperation with REUTERS.
[7] Payne (2003) and Hasbrouck and Saar (2004) also treat marketable limit orders as market orders.

ders. The latter is somewhat weakened for outstanding limit order *volume* and trading activity (as measured by number of trades per five minute interval), both of which tend to fall over the trading day. This should be due to the fact that trading at the UTS is not continuous so that traders enter limit orders rapidly when the market opens and do not submit large volumes towards the end of the trading session. Very similar intraday patterns are observed in a huge tick data sample for EBS trading in Tokyo by Ito and Hashimoto (2004).

Midquote returns, also displayed in Table 1, are mean zero, are heavily fat-tailed and show significant negative first order autocorrelation as documented in earlier microstructure analysis concerning FX and equities (see e.g. Payne, 2003, and Chung, van Ness and van Ness, 1999).

By and large, our data show similar characteristics and intraday patterns compared to currency trading in established markets, implying that insights gained from trading at the UTS might carry over to trading in other assets as well.

3.2 Construction of random trader groups and their characteristics

In order to systematically investigate determinants of price impacts, we have to cross-sectionally relate price impacts of different traders or trader groups to information determinants while controlling for possible inventory factors. Towards this end, we rely on the following randomization of trader groups. From the population of all 722 traders, we randomly assign each trader to one of two groups. The only restriction placed is that each of the so constructed groups consists of no more than 90% of all traders and not less than 10% of all traders. Therefore, we have two randomly formed trader groups which are also of random size. For example, the first group might represent 25% of all traders and the second group represents the remaining 75%. For each of the two groups we calculate the following six items:

1. the average trade size (ATS) of the group (i.e. mean transaction volume per trade),
2. the average market share of traders in this group (ARTS), i.e. the average of the market share (in terms of trading volume) of each trader belonging to this group,
3. the fraction of traders in the group that are located in a financial center (STFC),
4. the volume-weighted minute of the trading session (TT), the group trades at (measured in event time),
5. the average volume-weighted bid-ask spread (BAS) just prior to the trades of a respective group (we weight with the volumes of trades and the spreads are measured in event time) and

6. the average volume-weighted outstanding order book volume (OBV) just prior to the trades of the group (measured in event time).

Conceptually following Hasbrouck (1991a, b) we run a structural VAR with spot midquote returns, order flow from traders located in a financial center (FC) and order flow from traders outside the financial centers (NFC) as independent variables of the following form:

$$\mathrm{Ay}_{t+1} = \Gamma(\mathrm{L})\mathrm{y}_{t} + \mathrm{B}\upsilon_{t+1} \quad \text{with } \mathrm{Var}[\upsilon_{t+1}] = \mathrm{I}_3 \tag{1}$$

where

$$A = \begin{pmatrix} 1 & -\alpha_{12} & -\alpha_{12} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, y_t = \begin{pmatrix} r_t \\ x_t^{FC} \\ x_t^{NFC} \end{pmatrix} \; and \; B = \begin{pmatrix} \beta_r & 0 & 0 \\ 0 & \beta_{x^{FC}} & 0 \\ 0 & 0 & \beta_{x^{NFC}} \end{pmatrix} \tag{2}$$

and r denotes midquote spot returns, x^{FC} is order flow (in 100,000 USD) from traders located in Moskow or St. Petersburg and x^{NFC} is order flow from traders outside the two financial centers.[8] Γ(L) is a matrix polynomial in the lag operator and all three variables are measured on a frequency of one minute. This setup captures direct impacts of order flow on returns via α_{12} and α_{13} and the effect of past order flow via Γ(L). Compared to the original setup in Hasbrouck (1991a, b), we have added a second order flow variable which makes the system overidentified with one degree of freedom. The validity of this restriction can be tested. Since we are primarily interested in the total price impact per USD we compute the analytical long run price impact of order flow, Ξ_∞, by inverting the VAR operator[9] (see e.g. Lütkepohl, 2005)

$$\Xi_\infty = (I_3 - \Gamma_1 - \Gamma_2 - \ldots)^{-1} A^{-1} \tag{3}$$

Note that we do not multiply with the estimated standard errors of the system's innovations since we are interested in the price impact of order flow on a dollar-by-dollar basis. We then repeat this procedure 25,000 times and obtain a total of 50,000 price impacts and sets of group characteristics which we use to analyze the cross section of price impacts among artificial groups of traders.

Descriptive statistics for the so obtained price impacts and group characteristics can be found in Appendix 1. All variables are well behaved, i.e. have a sufficient range (which is the reason why we simulated such a large number of groups). Mean trade size for example ranges from 28,000 to 75,000 USD, the share of traders from a financial center covers groups from

[8] In Russia, the two political, economic and financial centers are Moskow and St. Petersburg, respectively. The remaining six regions are Ekaterinburg, N. Novgorod, Novosibirsk, Rostov, Samara and Vladivostock.

[9] The lag length of the VAR is determined by the SIC.

18% to more than 80% and the average trader size ranges from groups with traders who have an average market share of 0.04% (per trader) to groups with a share 0.29%. The average p-value for the χ^2-test of the overidentifying restriction in the SVAR in (1) and (2) is about 0.17 (not reported in the table) and thus is not rejected at a convenient level of significance.

Given these preliminary steps we are now able to relate price impacts to the six group characteristics. This allows us to draw inference about factors that influence the probability of informed trading.

4 Results

4.1 Bivariate relations

As a first step we look at simple bivariate correlations of price impacts and the six characteristics obtained for each trading group. Findings for our market are nicely in line with the literature.

Detailed results are shown in Table 2 and represent the correlation of our groups' six characteristics and their price impacts as well as correlations among the six characteristics. As it turns out, except for trading time, all other five items are positively correlated with the price impact. Consistent with economic intuition, groups of larger traders, groups with more traders from financial centers, groups that trade at higher spreads and groups that trade larger orders tend to have higher price impacts. Traders who place their orders earlier in the trading session have higher price impacts, confirming the experimental results of Bloomfield, O'Hara and Saar (2005). Somewhat astonishingly, groups that trade when the market is more liquid (as measured by outstanding order book volume) have higher price impacts. This might indicate that informed traders try to place their orders in times of high liquidity to lower their trading costs as discussed in Sections 1 and 2. This stands in sharp contrast to earlier results obtained in a time-series setting (see inter alia Hasbrouck and Seppi, 2001, Brandt and Kavajecz, 2004, or Payne, 2003).[10] However, it has to be kept in mind that these are bivariate relations in a cross section of trader groups so the earlier results from time series analyses are not contradicted by this finding. Whereas time series analyses focus on how the price impact of all trades varies between times of high or low liquidity, we look at how the average price impact of all trades from a certain group of traders varies according to their preferred level of liquidity. This is obviously not the same since our price impacts are measured over the whole sample space, whereas time-series study only looks at price impacts at certain times of the sample.

[10] In the time-series we also find lower price impacts when the market is more liquid, i.e. when outstanding order book volume is higher (not reported here). This confirms the results in Payne (2003).

Therefore, outstanding order book volume may well be positively correlated with price impacts if informed traders decide to trade when the market is more liquid but this result need not hold cross-sectionally once we control for information- related variables like trader size or the share of traders from financial centers. Therefore, a clearer picture might be obtained by measuring net effects of each variable on price impacts, which we turn to in the next section.

4.2 Weighted least squares regression analysis

We tackle the question of which traders' characteristics move prices by using weighted least-squares (WLS) regressions. The most important insight is quite conventional: signs and relative importance change in the multivariate approach.

We opt to use WLS instead of OLS since our data show heavy signs of heteroscedasticity. Furthermore, our price impacts are based on grouping traders which is known to induce residual heteroscedasticity (Greene, 2003). Therefore, we use the usual iterative procedure and first run an OLS regression, obtain the residuals and then determine a specific form of heteroscedasticity by regressing squared residuals on all six explanatory variables, their squares and the size of a respective trader group (which should be important due to the grouping of traders). Indeed, all variables contribute to the explanatory power and we use fitted squared residuals of this procedure to perform WLS regressions.

Table 3 shows results where the price impact of a group i (i =1, 2, …, 50,000) is regressed on the demeaned six characteristics and subsets of these.[11] Column (1) shows regression results when using all six items simultaneously. Except for trade size, which has a negative but insignificant coefficient estimate, all other five coefficient estimates show the expected signs and all coefficients are highly significant after controlling for heteroscedasticity. R^2s show that about 42% of the variation is explained by these six items. Controlling for trader size and share of informed traders indeed leads to a negative sign for outstanding order book volume, implying that this variable now accounts for liquidity effects.

The negative but insignificant coefficient on trade size is well in line with earlier work, which shows that the signed volume of a trade is not as important as the direction information alone (Bjønnes and Rime, 2005). However, in light of Chakravarty's findings on stealth trading for stock markets (Chakravarty, 2001) we include a squared trade size variable in the regression. Column (2) shows results when including all six variables plus squared trade size into the regression. Now, the squared trade size variable enters negatively, implying an in-

[11] We opt to demean the explanatory variables so that the estimated intercept (it actually is a "center-cept") gives the price impact for a typical trader group.

verted U-shaped influence of trade size on price impacts. The highest price impact implied by this estimate is obtained for a trade size of about 50,000 USD, which is, from Table 1, the average trade size in our sample. This hints at the validity of "stealth trading" for FX markets, too. Furthermore, in this specification, all three information variables, ATRS, STFC and TT keep their expected signs.

Columns (3) to (5) show results for specifications which successively eliminate all variables except for the two most clearly information-related variables, average trader size (ATRS) and share of traders from financial centers (STFC). As can be seen, our findings do not change when eliminating variables from the regressions and the absolute values of the estimated coefficients are similar in magnitude through all specifications. Finally, column (6) shows results for the full specification of column (2) when we exclude the 5,000 highest and lowest groups in terms of price impacts, i.e. we look at all trader groups with a price impact higher (lower) than the tenth (90^{th}) percentile of the empirical price impact distribution. This serves to eliminate the most extreme price impacts which might be obtained by having randomly sampled very extreme groups. It is obvious from the results in this restricted sample that our conclusions are not driven by outliers.

Apart from being statistically significant, the results are also of economic relevance. Based on the coefficient estimates of Table 3, we calculate the percentage change of the price impact for a one standard deviation increase in each of the six variables. Results of this exercise are shown in Table 4. Given an average midprice of about 31 RUR/USD and an average trade size of about 50,000 USD, the estimated intercept of 0.026 translates into an average price impact of about 4 pips. From Appendix 1 we have an average half-spread of roughly 8 pips, so that our base impact of 4 pips for an average trader (group) translates into a 50% share of the half spread. This indicates that the information component (or adverse selection costs) of the spread is roughly 50%, a magnitude much larger than that found in equity markets (Huang and Stoll, 1997) but very similar to those found for established FX markets (Payne, 2003, Bjønnes and Rime, 2005). This again highlights the fact that our market is quite similar compared to FX markets usually investigated in the literature.

As can be inferred from Table 4, both average trader size (ATS) and the share of traders from a financial center (STFC) have an economically significant effect on the price impact of about 5 to 9% and 23%, respectively. For the full specification corresponding to column (2), we also find the other four trader characteristics to be of clear economic importance.

As an interim finding we summarize: those traders' trades have the highest price impact who trade early in the day, when the spread is high, when the market is less liquid and when they are large traders from financial centers employing medium order sizes.

4.3 Simultaneous quantile regressions

Up to this point, we have only looked at the conditional mean of the price impact distribution. However, in this setting it is also interesting to examine the influence of traders' characteristics on price impacts at different points of the price impact distribution. We do, indeed, find non-linear relations and in particular we find a decreasing value of information. To provide an example: increasing the share of traders from financial centers in a group increases a group's price impact more sharply for groups that have small price impacts compared to groups that belong to a large quantile of the price impact distribution or vice versa.

To investigate these questions, we run simultaneous quantile regressions to obtain a more complete picture of the relationship of our six trader characteristics with the distribution of price impacts. More precisely, we simultaneously estimate quantile regressions at the 10, 25, 50, 75 and 90% quantile of the price impact distribution and obtain the variance covariance matrix estimate via bootstrapping. By estimating these nonparametric quantile regressions, we obtain a first glimpse at nonlinear effects and also produce estimates which are more robust to data outliers. As it turns out, the two most clearly information-related variables, trader size (ATRS) and share of traders from a financial center (STFC), are most important for the lower quantiles, whereas trading time (TT) and outstanding order book volume (OBV) change their signs when moving from low to higher quantiles.

We briefly sketch the methodology of quantile regressions more formally before we proceed with detailed results. Whereas the familiar least-squares estimators minimize a quadratic function

$$\min_{\beta \in \Re} \sum_i \left(y_i - x_i^T \beta\right)^2 \tag{4}$$

to model the conditional mean of the dependent variable, quantile regressions model the influence of explanatory variables on various quantiles of the distribution of the dependent variable and minimize a (possibly asymmetrically) weighted function of the data

$$\min_{\beta \in \Re} \sum_i \left(y_i - x_i^T \beta\right) \cdot \left|\alpha - I\left(y_i - x_i^T \beta < 0\right)\right| \tag{5}$$

where α is the quantile for which the estimation is performed and I(·) is the indicator function. Estimation may take place via simplex algorithms or GLS. It can be shown that this approach is robust to outliers, which follows intuitively from the fact that residuals do not enter squared

(Pagan and Ullah, 1999).[12] As indicated above, we perform this minimization simultaneously for the quantiles $\alpha = 0.1, 0.25, 0.5, 0.75, 0.9$ and use a bootstrap with 200 replications to obtain a covariance estimate which is used for all tests presented in the following.

Results of this estimation approach are presented in Table 5. F-tests for constancy of parameters for all quantiles are clearly rejected. Most notably, the influence of outstanding order book volume and the time of day variable change their signs when moving from the 10% to the 90% quantile. Since we are looking at a cross section of trader groups, care has to be taken about the interpretation of these results. Take, for example, the effect of outstanding order book volume, which is significantly positive in the lower quantiles and significantly negative in the 90% quantile. Therefore, adding traders who trade at higher order book volumes to a low-impact group (a group with a low price impact) increases the price impact of this group, given that it stays in the same quantile. This would be consistent with the information story of liquidity. Informed traders hit the market when there is lots of liquidity. However, adding traders who trade at higher order book volumes to a group which is already in the top decile of price impacts decreases the price impact of this group. Here, the liquidity effect seems to outweigh the additional information effect. The changing sign of the trading time variable shows that the negative effect of trading time on price impacts comes from the higher quantiles only, indicating the price impact distribution is contracted by later trading time.

While the effect of the spread seems to be at least numerically stable over the different quantiles, a larger trade size tends to decrease price impacts especially for high-impact groups. The remaining two information proxies, share of traders from a financial center and trader size, seem to unfold their positive influence mainly for low-impact groups, i.e. adding larger traders or more traders from a financial center to low-impact groups increases their price impact more sharply compared to groups that already belong to a high quantile of the price impact distribution. This makes intuitive sense and indicates that the relation of superior information and price impacts is concave rather than linear.

All in all, the results from the quantile regressions lead to similar conclusions but point at nonlinear relations between price impacts and traders' characteristics. Some of these nonlinearities, which are interesting from an economic perspective, are investigated in more detail in the next section.

4.4 Stealth trading and combined effects of trader characteristics

[12] Clearly, the 50% quantile regression minimizes the absolute residuals and yields the LAD estimator (Koenker and Bassett, 1978).

This section examines the price impact of one of the most prominent indicators of informed trading, i.e. trade size, in more detail. We find from kernel regressions that medium sized trades indeed have the highest price impact, which holds for large traders and traders from financial centers, too. This confirms the "stealth trading" hypothesis for foreign exchange and extends it to further trader characteristics.

As noted in the introduction, there are numerous empirical and theoretical papers relating the information contained in the trading process to observable market statistics. The most important such relation in the empirical literature may be trade size. Empirically, the relation between signed trading volume and price impacts is less clear. Most researchers find signed trade indicators to yield better explanatory power when explaining price changes (Bjønnes and Rime, 2005). Furthermore, Chakravarty (2001) finds, for a sample of NYSE stocks, that the relation between trade size and price changes is non-linear and also depends on the initiator of the trades. Medium sized trades of institutional investors have the highest price impacts, which is affirmative of the so-called "stealth trading" hypothesis of Barclay and Warner (1993), which holds that informed traders choose medium-sized orders to avoid giving away their information too easily (by trading very large orders which save transaction costs) and to avoid excessive trading costs (by trading very small orders that hide their information). Our data and empirical approach are well suited to address this issue for FX markets and to extend the analysis of Chakravarty (2001) to continuous measures of information proxies.

The results on the effect of squared trade size in Table 3 already suggested that traders who tend to place medium-sized trades have the highest average price impact for all their trades. This section considers a more general setup, i.e. nonparametric kernel regressions, to shed light on the influence of trade size in combination with other characteristics on the price impact of a group.[13] This is because Chakravarty (2001) not only finds medium-size trades to be most informative but rather trades of medium size that originate from institutional investors. A close analogue to this finding in our data set would be that traders who tend to place medium-sized trades and who are large or from a financial center have the highest price impact on average.

As mentioned above, we use kernel regressions of the form

$$PI_i = g(X_i) + \xi_i \qquad (6)$$

where PI_i is group i's price impact and X_i contains all six characteristics for the same group. The functional form of $g(\cdot)$ is left unspecified so that the conditional expectation $E[Y_i|X_i]$

[13] Kernel regressions have become quite common in finance, see e.g. Aït-Sahalia and Lo (1998) or Evans and Lyons (2002).

cannot be misspecified in the usual sense (Pagan and Ullah, 1999). This is clearly desirable here, since we do not want to impose any functional restrictions on the relationship between price impacts, trade size and other group characteristics.

Estimation of the fitted price impact dependent on a particular value of the explanatory variables x_0 is done the standard way by calculating

$$\hat{g}(x_0;x)=\frac{\sum_{i=1}^{n} y_i K\left(\frac{x_0-x_{i\bullet}^T}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x_0-x_{i\bullet}^T}{h}\right)} \tag{7}$$

where h is a bandwidth parameter, K(·) is a Gaussian product kernel and i=1,…,50,000 denotes observations and equation (7) is the well-known Nadaraya-Watson (NW) kernel regression estimator (see Pagan and Ullah, 1999). The bandwidth h controls the smoothness of the fit and is usually chosen optimally to trade-off bias and efficiency. If standardized explanatory variables are used (i.e. standardized to have unit variances) it can be shown that the following bandwidth

$$h_* = \left(\frac{4}{m+2}\right)^{\frac{1}{m+4}} n^{-\frac{1}{m+4}} \tag{8}$$

is optimal for the Gaussian product kernel where in our case m=4 and n=50,000 and $h_*\approx 0.32$. In the following analysis we show results for h=0.4 which yields qualitatively identical but somewhat smoother results which are better suited for our graphical analysis.[14]

We now use this approach to take a look at the interaction of trader size, share of informed traders, trade size and price impacts by computing the expected price impact according to equation (7), while holding fixed all explanatory variables at their unconditional mean except for two variables we are interested in as detailed in the analysis below.

Figure 2, Panel A, shows a surface plot of price impacts as a function of trade size (horizontal axis) and the share of informed traders (vertical axis), i.e. we set all except these two variables at their unconditional mean and estimate (7) for different combinations of ATS and STFC. In Figure 2, darker areas mark higher price impacts. As can be seen, price impacts generally increase when adding more traders from a financial center to a group. Furthermore, long-run cumulative price impacts are highest for medium-sized trades from FC traders, not

[14] Since we are dealing with six explanatory variables the sample size of 50,000 observations is actually quite small. We have also employed semi-parametric approaches (not reported here) where the spread, trading time and outstanding order book volume are restricted to enter the regression linearly so that only trade size, trader size and share of traders from financial centers enter the kernel regression part. The results were virtually identical.

for small or large trade sizes, which confirms the stealth trading hypothesis as discussed above for FX markets for the first time.

Panel B of the same figure shows price impacts as function of trader size (vertical axis) and trade size (horizontal axis). Again, small and medium sized orders from large traders are most informative, confirming the stealth trading hypothesis.

Finally, Panel C shows results when looking at price impacts depending on the share of FC traders and average trader size, now holding fixed trade size at its mean. As can be expected from the above discussion, price impacts are highest for large traders from financial centers. Interestingly, the level curves of price impacts have a convex shape. One may be tempted to draw an analogy to conventional production functions of the Cobb-Douglas type. In our case, information is the commodity to be produced, whereas trader size and the share of traders from financial centers are the production factors: then trader size and location "produce" information with marginally declining returns.

5 Robustness analyses

In this section, we perform some robustness analyses which include an investigation of possible multicollinearity, tests on two separate subsamples of our data and an adjustment of order flow for autocorrelation in individual order flow. Results are robust with respect to these modifications.

It might be argued that our explanatory variables are not uncorrelated, thereby harming the validity of the results obtained so far. Specifically, multicollinearity of our trading characteristics might be an issue. Large traders are presumably from financial centers, and both should trade at higher order book volumes and earlier in the day if the information story presented above is true. This is surely an issue, and we find clear evidence for this. The correlation between trader size and the share of FC traders is about 0.30, the correlation of outstanding order book volume and share of FC traders (trader size) is 0.47 (0.30). As can be expected, the trading time variable and outstanding order book volume correlate with a coefficient of -0.66. As is well known, the only feasible way in empirical applications to overcome multicollinearity is to "use more information" in the form of imposing priors on the coefficients or by obtaining more data (Greene, 2003). However, increasing or decreasing the number of simulated groups by some 10,000 does not change the economic or statistical significance of our results. Furthermore, as the results in Table 3 show, using only a subset of our variables leads to similar results as in the full specification. In addition, due to our large cross-section, commonly used measures of the disturbing effect of multicollinearity, Variance inflation factors (VIFs) and condition numbers do not indicate a problem, as can be seen in Appendix 2.[15]

As a second sensivity check, we rerun the generation of random trader groups and simulate two separate cross-sections, one for the first five days of our data set and a second for the remaining four days. The results are not significantly different to those obtained from using the whole sample space so we do not report results here.

Since we are interested in the information content of trades, it is necessary to estimate the effect or order flow *innovations* on midquote returns. Observed order flows may be a flawed measure of new information if traders use splitting order strategies (see e.g. Bernhardt and Hughson, 1997 or Chordia and Subrahmanyam, 2004), i.e. traders might split otherwise large orders into smaller portions to hide their trading intentions. This being true, we should observe significant autocorrelation of individual traders' order flows on given days. In order to address this issue, we also estimate total price impacts for the same 25,000 repetitions with an "adjusted" order flow measure. This is constructed by adjusting individual traders' order flow for up-to second order. More precisely, for each trader *i* and each day *d,* we estimate an AR(2)-model and use the innovation of the AR(2) model as the adjusted order flow,

$$x_k^{adj,i,d} = x_k^{i,d} - \hat{\alpha}_1^{i,d} x_{k-1}^{i,d} - \hat{\alpha}_2^{i,d} x_{k-2}^{i,d} \qquad (9)$$

where $x^{i,d}$ denotes order flow of trader *i* and day *d*, *k* is the event time index on the tick-by-tick data set and, of course, the first two trades of trader *i* and day *d* cannot be adjusted.[16] As it turns out, there is autocorrelation in individual order flow. Although the average estimated autoregressive parameters are near zero with a moderate R^2 of 12%, there are a lot of cases with large R^2s and AR coefficients. For example, looking only at the 257 positive coefficient estimates of α_1 yields an average estimate of 0.34 (0.39 for α_2). A similar picture emerges for the remaining 132 negative estimates of α_1 which average -0.33 (-0.37 for α_2). In short, there are a lot of cases with significant autocorrelation that point towards possible splitting order strategies. However, all results reported in the preceding sections are virtually unchanged when we use adjusted order flow instead of the usual order flow measure, so we do not report the results.[17]

---

[15] A commonly used rule of thumb is that VIFs should be smaller than 10 und condition numbers should not exceed 10. Our values are far below these numbers.

[16] We do this adjustment for each trader who has at least ten trades on a given day.

6 Conclusions

Financial markets are a means to aggregate information that is widespread in the economy. Order flow may transport information of asymmetrically-informed market participants. Accordingly, it has been shown in theoretical models and empirical studies alike that order flow has a permanent impact on prices. Credibility of this information story could be increased by demonstrating the information aggregation process in more detail. One important step in this direction seems to be identifying those traders who bring information into the market, which obviously requires trading data being linked to trader identities. In this respect, our data go beyond available material – according to the best of our knowledge – and thus allow for analyses that have not been performed before.

Due to our focus on identifying whose traders' trades convey more or less information, we generate 50,000 groups whose price impacts can be related to their trading behavior and their likely information endowment. Equipped with this data, we perform a cross-sectional analysis where we regress the price impacts of (randomly formed) trader groups on six indicators of information.

Our major finding is the determination of six *trader's* trade characteristics – characteristics that theory has interpreted as indicators of information – that help to identify informed *trading*. It is shown in the multivariate approach that traders who convey more information, i.e. technically whose market orders have a larger permanent price impact, use medium-sized orders, have a large trading volume, are located in a financial center, trade early in the trading session, trade when spread is high and when the order book is relatively thin. We show in quantile regressions that the effect from these trade characteristic variables is non-linear in the sense of a declining marginal importance. Finally, findings are robust to several modifications, such as splitting the sample or using a refined order flow measure.

As a focus of our research, we examine the effect of trade size as an indicator of informed trade with particular scrutiny. It is also a good example to show that our cross-sectional analysis of traders can bring about different results compared to other approaches. If one relates trade size to price impact (the most conventional approach), there is no relation in our limit order market. If one relates trade size to traders, however, one finds that informed traders tend to use larger trade sizes. It is revealing in this respect that this picture changes again if one examines the same relation controlling for further indicators of informed trade. Then the relation becomes non-linear, as traders who convey most information use medium-

[17] Results are available from the authors upon request.

sized trades. This finding of so-called stealth trading (see Barclay and Warner, 1993, Chakravarty, 2001) is robust to a detailed examination.

This paper shows that the conveyance of information via *trades* varies markedly across different *traders*. Further research might be interested in extending the cross-sectional analysis by examining the interaction between different kinds of traders and, therefore, how the initial information of these informed traders actually becomes embedded in prices.

References

Admati, Anat R. and Paul Pfleiderer (1988), A Theory of Intraday Trading Patterns: Volume and Price Variability, *Review of Financial Studies*, 1, 3-40.

Aït-Sahalia, Yacine and Andrew W. Lo (1998), Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, *Journal of Finance*, 53:2, 499-547.

Barclay, Michael J. and Jerold B. Warner (1993), Stealth Trading and Volatility: Which Trades Move Prices?, *Journal of Financial Economics*, 34:3, 281-306.

Bernhardt, Dan and Eric Hughson (1997), Splitting Orders, *Review of Financial Studies*, 10:1, 69-101.

Biais, Bruno, Pierre Hillion and Chester S. Spatt (1995), An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse, *Journal of Finance*, 50:5, 1655-1689.

Bjønnes, Geir H. and Dagfinn Rime (2005), Dealer Behavior and Trading Systems in Foreign Exchange Markets, *Journal of Financial Economics*, 75:3, 571-605.

Bjønnes, Geir H., Dagfinn Rime and Haakon O. Solheim (2004), Liquidity Provision in the Overnight Foreign Exchange Market*, Journal of International Money and Finance*, 24:2, 175-196.

Bloomfield, Robert, Maureen O'Hara and Gideon Saar (2005), The 'Make or Take' Decision in an Electronic Market: Evidence on the Evolution of Liquidity, *Journal of Financial Economics*, 75, 165-199.

Brandt, Michael W. and Kenneth A. Kavajecz (2004), Price Discovery in the U.S. Treasury Market: The Impact of Orderflow and Liquidity on the Yield Curve, *Journal of Finance*, 59:6, 2623-2654.

Campbell, John Y., Tarun Ramadorai and Tuomo Vuolteenaho (2005), Caught On Tape: Institutional Order Flow and Stock Returns, NBER Working Paper 11439.

Chakravarty, Sugato (2001), Stealth-trading: Which Trader's Trades Move Stock Prices?, *Journal of Financial Economics*, 61:2, 289-307.

Cheung, Yin-Wong and Menzie D. Chinn (2001), Currency Traders and Exchange Rate Dynamics: A Survey of the US Market, *Journal of International Money and Finance*, 20, 439-471.

Chordia, Tarun and Avanidhar Subrahmanyam (2004), Order Imbalance and Individual Stock Returns: Theory and Evidence, *Journal of Financial Economics*, 72:3, 485-518.

Chung, Kee H., Bonnie F. Van Ness and Robert A. Van Ness (1999), Limit Orders and the Bid-Ask Spread, *Journal of Financial Economics*, 53, 255-287.

Copeland, Thomas E. and Dan Galai (1983), Information Effects on the Bid-Ask Spread, *Journal of Finance*, 38, 1457-1469.

Coval, Joshua D. and Tobias J. Moskowitz (2001), The Geography of Investment: Informed Trading and Asset Prices, *Journal of Political Economy*, 109:4, 811-841.

Covrig, Vicentiu and Michael Melvin (2002), Asymmetric Information and Price Discovery in the FX Market: Does Tokyo Know more about the Yen?. *Journal of Empirical Finance*, 9, 271-285.

Demsetz, Harold (1968), The Cost of Transacting, *Quarterly Journal of Economics*, 82, 33-53.

Dufour, Alfonso and Robert F. Engle (2001), Time and the Price Impact of a Trade, *Journal of Finance*, 55:6, 2467-2498.

Dunne, Peter, Harald Hau and Michael Moore (2004), Macroeconomic Order Flows: Explaining Equity and Exchange Rate Returns, CEPR Discussion Paper 4806.

Easley, David and Maureen O'Hara (1987), Price, Trade Size and Information in Securities Markets, *Journal of Financial Economics*, 19, 69-90.

Evans, Martin D. (2002), FX Trading and Exchange Rate Dynamics, *Journal of Finance*, 57:6, 2405-2447.

Evans, Martin D.D. and Richard K. Lyons (2002), Order Flow and Exchange Rate Dynamics, *Journal of Political Economy*, 110, 170-180.

Evans, Martin D. and Richard K. Lyons (2002a), Time-Varying Liquidity in Foreign Exchange, *Journal of Monetary Economics*, 49, 1025-1051.

Evans, Martin D. and Richard K. Lyons (2005), Meese-Rogoff Redux: Micro-Based Exchange Rate Forecasting, *American Economic Review*, 95:2, 405-414.

Evans, Martin D. and Richard K. Lyons (2005a), Exchange Rate Fundamentals and Order Flow, Working paper, Haas School of Business, U.C. Berkeley.

Froot, Kenneth A. and Tarun Ramadorai (2005), Currency Returns, Institutional Investor Flows and Exchange Rate Fundamentals, *Journal of Finance,* 60:3, 1535-1565.

Gehrig, Thomas and Lukas Menkhoff (2004), The Use of Flow Analysis in Foreign Exchange: Exploratory Evidence, *Journal of International Money and Finance*, 23, 573-594.

Glosten, Lawrence R. (1994), Is the Electronic Open Limit Order Book Inevitable?, *Journal of Finance*, 49:4, 1127-1161.

Glosten, Lawrence R. and Paul R. Milgrom (1985), Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, *Journal of Financial Economics*, 14, 71–100.

Goldberg, Linda S. and Rafael Tenorio (1997), Strategic Trading in a Two-sided Foreign Exchange Auction, *Journal of International Economics*, 42, 299-326.

Green, T. Clifton (2004), Economic News and the Impact of Trading on Bond Prices, *Journal of Finance*, 59:3, 1201-1234.

Greene, William H. (2003), *Econometric Analysis*, New Jersey: Prentice Hall, 5.ed.

Hasbrouck, Joel (1991), Measuring the Information Content of Stock Trades, *Journal of Finance*, 46:1, 179-207.

Hasbrouck, Joel (1991a), The Summary Informativeness of Stock Trades: An Econometric Analysis, *Review of Financial Studies*, 4:3, 571-595.

Hasbrouck, Joel (2006), *Empirical Market Microstructure*, Oxford et al.: Oxford University Press.

Hasbrouck, Joel and Gideon Saar (2004), Technology and Liquidity Provision: The Blurring of Traditional Definitions, Working Paper, Stern School of Business, New York University.

Hasbrouck, Joel and Duane J. Seppi (2001), Common Factors in Prices, Order Flows, and Liquidity, *Journal of Financial Economics*, 59, 383-411.

Hau, Harald (2001), Location Matters: An Examination of Trading Profits, *Journal of Finance*, 56:3, 1959-1983.

Ho, Thomas and Hans Stoll (1981), Optimal Dealer Pricing Under Transactions and Return Uncertainty, *Journal of Financial Economics*, 9:1, 47-73

Hollifield, Burton, Robert A. Miller and Patrik Sandås (2003), Empirical Analysis of Limit Order Markets, *Review of Economic Studies*, 71, 1027-1063.

Huang, Roger D. and Hans R. Stoll (1994), Market Microstructure and Stock Return Predictions, *Review of Financial Studies*, 7, 179–213.

Ito, Takatoshi and Yuko Hashimoto (2004), Microstructure of the Yen/Dollar Foreign Exchange Market: Patterns of Intra-Day Activity Revealed in the Electronic Broking System, NBER Working Paper 10856.

Ito, Takatoshi, Richard K. Lyons and Michael Melvin (1998), Is There Private Information in the FX Market? The Tokyo Experiment, *Journal of Finance*, 53:3, 1111-1130.

Jones, Charles M., and Marc L. Lipson (2004), Are Retail Orders Different?, Working Paper, Columbia University.

Koenker, Roger and Gilbert Bassett (1978), Regression Quantiles, *Econometrica*, 46:1, 33-50.

Koski, Jennifer L. and Roni Michaely (2000), Prices, Liquidity, and the Information Content of Trades, *Review of Financial Studies*, 13:3, 659-696.

Kyle, Albert S. (1985), Continuous Auctions and Insider Trading, *Econometrica*, 53, 1315-1336.

Lütkepohl, Helmut (2005), *New Introduction to Multiple Time Series Analysis*, Berlin et al.: Springer.

Lyons, Richard K. (2001), *The Microstructure Approach to Exchange Rates*, Cambridge: MIT Press.

Madhavan, Ananth and Seymour Smidt (1991), A Bayesian Model of Intraday Specialist Pricing, *Journal of Financial Economics*, 30, 99-134.

Malloy, Christopher (2005), The Geography of Equity Analysis, *Journal of Finance*, 60:2, 719-756.

Marsh, Ian W. and Ceire O'Rourke (2005), Customer Order Flow and Exchange Rate Movements: Is there Really Information Content?, Working Paper, Cass Business School London.

Osler, Carol, Alexander Mende and Lukas Menkhoff (2006), Price Discovery in Currency Markets, Working Paper, Brandeis University.

Pagan, Adrian and Aman Ullah (1999), *Nonparametric Econometrics*, Cambridge: Cambridge University Press.

Payne, Richard (2003), Informed Trade in Spot Foreign Exchange Markets: An Empirical Investigation, *Journal of International Economics*, 61, 307-329.

Peiers, Bettina (1997), Informed Traders, Intervention, and Price Leadership: A Deeper View of the Microstructure of the Foreign Exchange Market, *Journal of Finance*, 52:4, 1589-1614.

Sias, Richard and Laura T. Starks (1997), Return Autocorrelation and Institutional Investors, *Journal of Financial Economics*, 46:1, 103-131.

Sias, Richard, Laura T. Starks and Sheridan Titman (2004), Changes in Institutional Ownership and Stock Returns: Assessment and Methodology, *Journal of Business*, forthcoming.

Table 1. Descriptive statistics

This table shows descriptive statistics for RUR/USD spot returns and the evolution of the limit order book for the whole sample period (all) and for non-overlapping five minute subsamples (rows "5" to "60"), where "5" denotes the first five minutes of the trading sessions, "10" denotes minutes five to ten of the trading sessions and so on. Columns two to six show the first four moments of the return distribution and first order autocorrelation of returns (ρ_{-1}). OBV is outstanding order book volume in mill. USD whereas OBC shows the number of outstanding orders. Trade size denotes the average size of a market order in USD and "no. trades" shows the number of market orders for a given sample. The last column shows the percentage spread.

	mean	st. dev.	skewness	Kurtosis	ρ_{-1}	OBV	OBC	trade size	no. trades	pct. spread
all	0.000002	0.000301	-0.0802	18.70	-0.0961	17.61	165.96	49,396	14,109	0.0071
5	-0.000001	0.000276	-0.4299	24.65	-0.1318	19.81	145.35	55,795	3,140	0.0115
10	0.000001	0.000294	-0.0955	18.80	-0.1070	22.86	204.63	52,236	2,404	0.0045
15	0.000013	0.000289	0.3732	16.32	-0.1361	19.18	191.60	49,009	1,907	0.0043
20	-0.000003	0.000290	-0.3718	18.82	-0.0600	18.39	187.51	47,362	1,242	0.0049
25	-0.000009	0.000299	-0.3256	19.03	-0.0447	15.70	184.41	46,821	1,024	0.0049
30	-0.000011	0.000308	-0.0777	17.02	-0.0132	15.86	173.98	39,200	832	0.0046
35	0.000004	0.000321	-0.0760	15.96	-0.1488	15.00	171.48	44,903	585	0.0050
40	0.000005	0.000287	0.1154	18.89	-0.5050	13.48	152.89	50,000	760	0.0049
45	0.000004	0.000352	0.0060	14.09	-0.0895	11.56	136.63	51,427	597	0.0045
50	0.000018	0.000345	0.2554	16.26	-0.2230	10.58	129.88	42,732	541	0.0045
55	0.000018	0.000358	0.3254	13.32	-0.0459	9.92	98.59	39,900	581	0.0059
60	-0.000002	0.000324	-0.3516	18.35	-0.1420	10.58	71.93	44,429	496	0.0120

Figure 1. RUR/USD spot exchange rate

This figure shows the evolution of the spot RUR/USD (vertical axis) over the nine trading days of our sample. The figure is based on midquotes in event time (all trades).

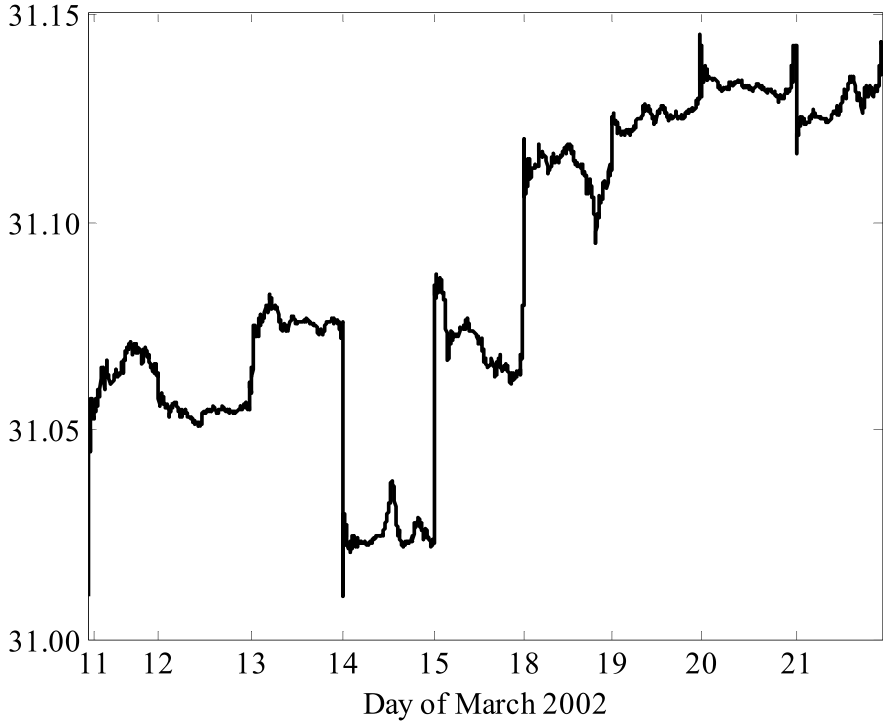


Table 2. Correlation of price impacts and trader group characteristics

This table shows correlation coefficients of price impacts (PI) and all six trader group characteristics: (average) trade size in USD (ATS ($\times 10^{-5}$)), (average) trader size (ATRS) calculated as the average market share of traders in a respective group (in %), share of traders from a financial center (STFC), trading time (TT), bid-ask spread (BAS), and outstanding order book volume (OBV).

	PI	ATS	ATRS	STFC	TT	BAS
ATS	0.22					
ATRS	0.26	0.41				
STFC	0.28	0.17	0.30			
TT	-0.24	-0.12	-0.08	-0.37		
BAS	0.59	0.34	0.16	0.07	-0.19	
OBV	0.13	0.16	0.15	0.48	-0.66	-0.05

Table 3. Weighted least squares regression results

This table shows results from WLS regressions of a group's long run price impact (calculated according to equation (2) and (3)) on the trader group characteristics: (average) trade size in USD (ATS ($\times 10^{-5}$)), (average) trader size (ATRS) calculated as the average market share of traders in a respective group (in %), share of traders from a financial center (STFC), trading time (TT), bid-ask spread (BAS), and outstanding order book volume (OBV). All regression results are based on White's (1980) heteroscedasticity consistent standard errors. Numbers in parentheses denote t-values for coefficient estimates and p-values for White's heteroscedasticity test in the last row (Het test).

	(1)	(2)	(3)	(4)	(5)	(6)
ATS	-0.015	0.026	0.135			0.042
	(-0.90)	(1.47)	(5.56)			(2.22)
ATS^2		-1.230	-1.748			-1.354
		(-3.28)	(-13.24)			(-3.21)
ATRS	0.064	0.051	0.046	0.086	0.084	0.053
	(4.15)	(3.46)	(2.34)	(4.83)	(4.75)	(3.83)
STFC	0.133	0.129	0.137	0.154	0.168	0.128
	(5.67)	(6.33)	(4.97)	(4.07)	(4.53)	(6.35)
TT	-0.003	-0.002	-0.002	-0.002		-0.002
	(10.66)	(-2.86)	(-2.25)	(-2.69)		(-2.62)
BAS	0.003	0.003				0.003
	(10.66)	(13.36)				(9.58)
OBV	-0.021	-0.017				-0.019
	(-3.89)	(-4.48)				(-4.99)
Const.	0.026	0.026	0.026	0.026		0.026
	(50.41)	(49.92)	(31.21)	(26.05)		(53.72)
adj. R^2	0.42	0.43	0.15	0.13	0.11	0.42
AIC	-6.02	-6.07	-5.59	-5.33	-5.32	-6.41
Het test	37.66	55.86	43.38	28.10	25.64	38.46
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
obs	50,000	50,000	50,000	50,000	50,000	40,000

Table 4. Economic significance

This table shows the effect of a one standard deviation increase in one of the six trader group characteristics on the percentage deviation of price impacts from their conditional mean while holding all other five characteristics fixed. The five columns show the effect on price impacts according to the specifications in Table 3. The base impact is reported in pips and is calculated by setting all variables to their unconditional mean and employing a midprice of 31.00 RUR/USD and an average order size of 50,000 USD.

	(1)	(2)	(3)	(4)	(5)	(6)
Base impact (pips)	3.98	4.04	4.10	4.09	4.07	4.03
ATS	-2.22%	-11.51%	-1.53%			-10.72%
ATRS	7.07%	5.55%	4.93%	9.36%	9.23%	5.83%
STFC	23.54%	22.57%	23.61%	26.71%	29.11%	22.46%
TT	-12.39%	-8.14%	-5.57%	-7.38%		-7.20%
BAS	32.11%	32.46%				27.97%
OBV	-21.83%	-17.23%				-19.24%

Table 5. Simultaneous quantile regressions

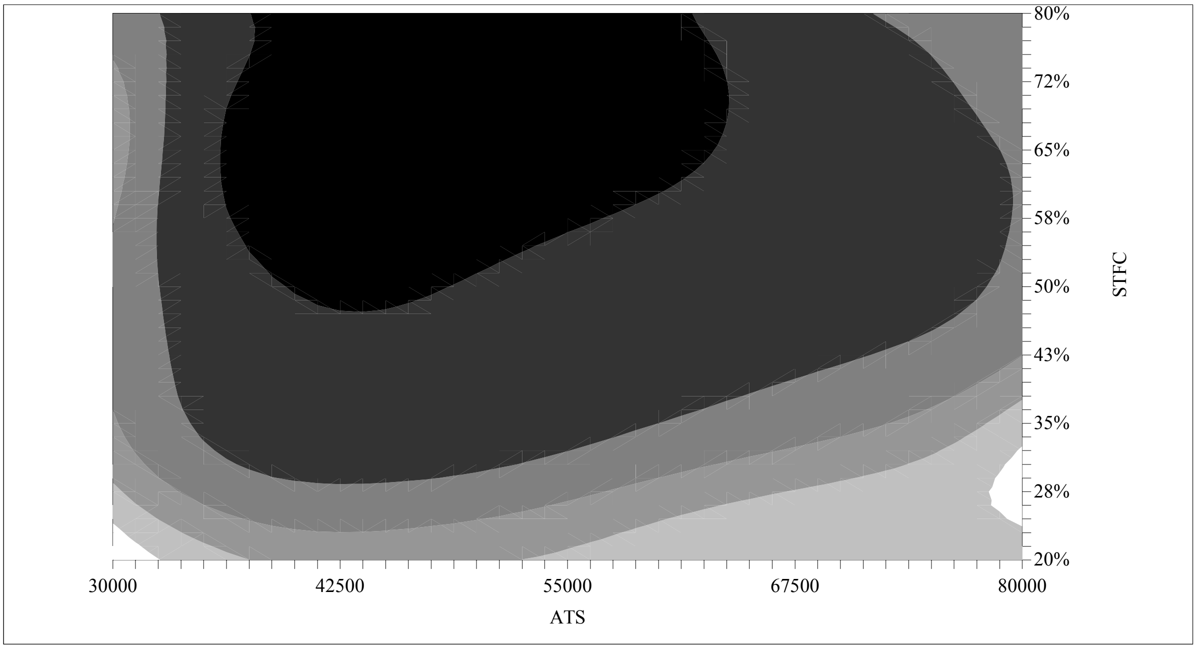
This table shows results for simultaneous quantile regressions from equation (5). Included variables are: (average) trade size in USD (ATS ($\times 10^{-5}$)), (average) trader size (ATRS) calculated as the average market share of traders in a respective group (in %), share of traders from a financial center (STFC), trading time (TT), bid-ask spread (BAS), and outstanding order book volume (OBV). The last column shows F-tests for constancy of parameters across quantiles and numbers in parentheses are t-values for coefficient estimates and p-values for the F-tests. Inference is based on a bootstrap with 200 replications.

	α=0.10	α=0.25	α=0.50	α=0.75	α=0.90	F
ATS	-0.030	-0.037	-0.039	-0.036	-0.04	2.83
	(-10.93)	(-16.57)	(-22.86)	(-21.21)	(-15.00)	(0.00)
ATRS	0.168	0.133	0.090	0.046	0.010	154.31
	(26.08)	(30.61)	(25.64)	(10.76)	(1.79)	(0.00)
STFC	0.070	0.063	0.055	0.050	0.040	30.35
	(24.57)	(33.11)	(29.90)	(24.55)	(14.68)	(0.00)
TT	0.010	-0.000	-0.001	-0.001	-0.002	48.61
	(6.56)	(-1.44)	(-4.83)	(-6.81)	(-9.60)	(0.00)
BAS	0.004	0.004	0.004	0.004	0.004	20.41
	(97.33)	(126.73)	(150.39)	(134.03)	(86.07)	(0.00)
OBV	0.007	0.003	-0.001	-0.000	-0.002	33.05
	(12.05)	(7.61)	(-3.00)	(-0.26)	(-3.59)	(0.00)
constant	0.011	0.018	0.024	0.029	0.035	
	(91.29)	(285.36)	(642.63)	(633.83)	(371.59)	
Pseudo R^2	0.25	0.31	0.32	0.30	0.28	

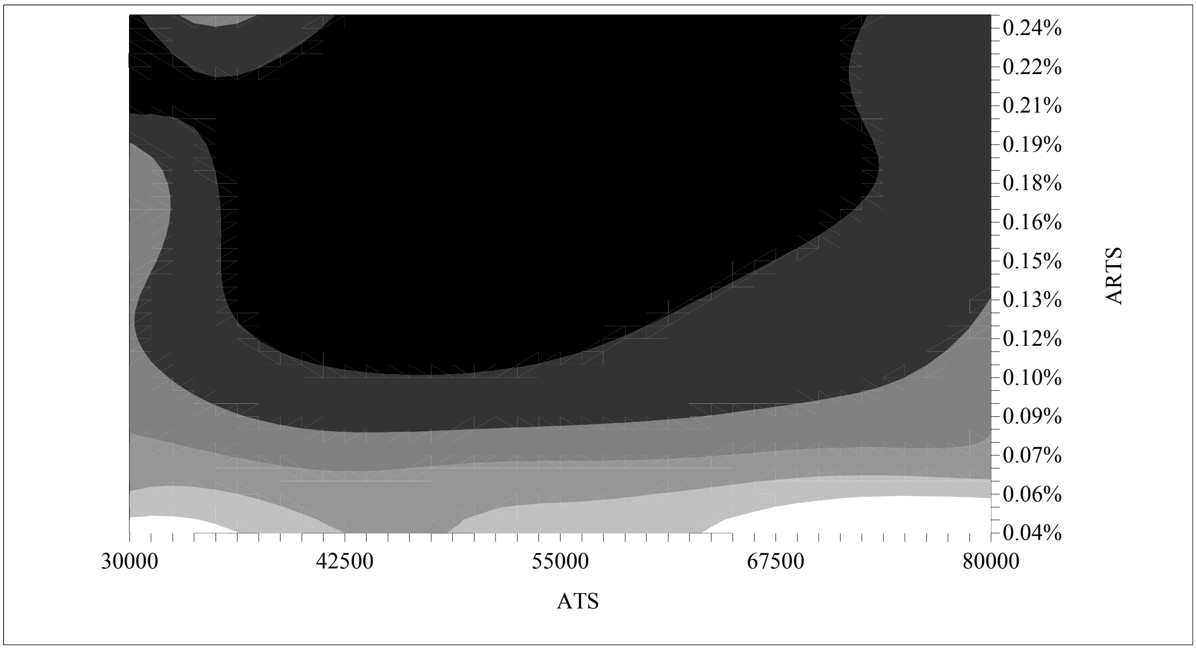
Figure 2. Kernel regression: price impacts, trade size, FC traders and trader size

This figure shows fitted price impacts obtained from nonparametric kernel regressions and depending on two of the six group characteristics while all other four characteristics are held fixed at their unconditional mean. Darker areas indicate higher price impacts. Panel (A) shows price impacts as a function of average trade size (ATS) and the share of traders from a financial center (STFC). Panel (B) shows price impacts depending on average trade size and average trader size (ATRS). Panel (C) shows price impacts for different combinations of the share of traders from a financial center and average trader size.

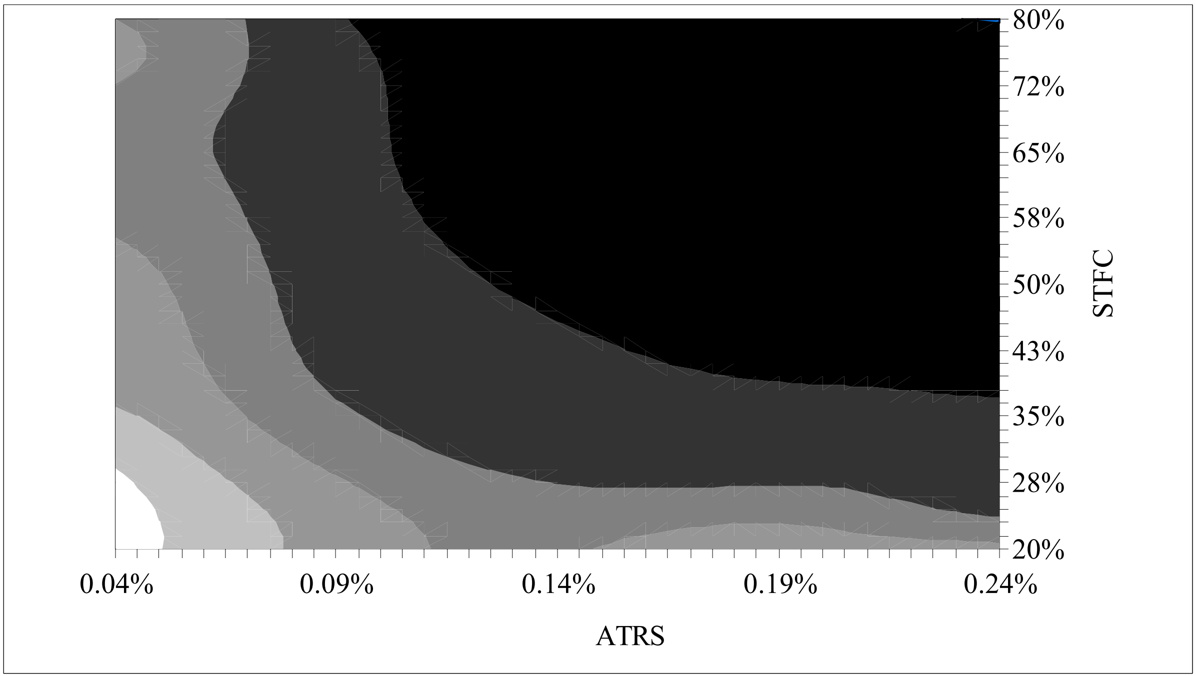
Panel A: STFC and ATS



Panel B: ARTS and ATS



Panel C: STFC and ARTS



Appendix 1. Descriptive statistics for simulated trader groups

This table shows descripticve statistics for price impacts (PI) and the six trader group characteristics: (average) trade size in USD (ATS ($\times 10^{-5}$)), (average) trader size (ATRS) calculated as the average market share of traders in a respective group (in %), share of traders from a financial center (STFC), trading time (TT), bid-ask spread (BAS), and outstanding order book volume (OBV). q(·) denotes the quantile for the value in brackets.

	PI	ATS	ATRS	STFC	TT	BAS	OBV
mean	0.023	0.497	0.138	0.638	18.629	16.628	17.972
max.	0.195	0.746	0.289	0.801	26.939	38.326	20.159
min.	-0.141	0.289	0.039	0.181	10.114	9.708	15.638
q(10)	0.005	0.452	0.117	0.588	17.597	14.070	17.682
q(25)	0.015	0.476	0.129	0.619	18.162	14.880	17.853
q(50)	0.025	0.496	0.138	0.640	18.616	16.681	17.979
q(75)	0.031	0.517	0.147	0.662	19.054	17.705	18.101
q(90)	0.040	0.543	0.158	0.688	19.646	19.252	18.254

Appendix 2. Multicollinearity analysis

This table shows variance inflation factors (VIF) and the condition number for the regression analysis of Table 3, column (1). Variables are: (average) trade size in USD (ATS ($\times 10^{-5}$)), (average) trader size (ATRS) calculated as the average market share of traders in a respective group (in %), share of traders from a financial center (STFC), trading time (TT), bid-ask spread (BAS), and outstanding order book volume (OBV). Variance inflation factors are calculated as $1/(1 - R^2)$, where R^2 is the R-squared in a regression of each independent variable on the remaining independent variables.

	VIF
ATS	1.35
ATRS	1.30
STFC	1.41
TT	1.97
BAS	1.27
OBV	2.17
Condition number	2.88