Whose trades convey information?
Evidence from a cross-section of traders

Lukas Menkhoff and Maik Schmeling,
Leibniz Universität Hannover, Germany

Abstract:
This paper contributes empirically to our understanding of informed traders. It analyzes *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.

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

February 6, 2007

We would like to thank participants at European Science Fund-workshop on microstructure in Warwick, at the European Economic Association conference in Vienna, at several university seminars and in particular Thomas Gehrig, Thomas Lux, Carol Osler and David Veredas. We gratefully acknowledge research assistance by Leila Gadijeva and financial support by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG).

Lukas Menkhoff, Department of Economics, Leibniz Universität Hannover, Königsworther Platz 1, D-30167 Hannover, Germany, menkhoff@gif.uni-hannover.de;
Maik Schmeling, Department of Economics, Leibniz Universität Hannover, Königsworther Platz 1, D-30167 Hannover, Germany, schmeling@gif.uni-hannover.de

Whose trades convey information?
Evidence from a cross-section of traders

Abstract:

This paper contributes empirically to our understanding of informed traders. It analyzes *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.

Whose trades convey information? Evidence from a cross-section of traders

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, 2007). 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.

Theory has indeed provided guidance which kind of traders might be informed (see Section 2). However, despite its core importance and seeming simplicity, the empirical identification of informed traders is still listed among the big open questions in the microstructure literature (Lyons, 2001, Hasbrouck, 2007). 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 trader's trades in the whole market (segment) including every single characteristic of these trades.[2] In reality, however, there is a shortage of data and 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

[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, Bozcuk and Lasfer (2005), Brandt and Kavajecz (2004), Green (2004) and Hautsch and Hess (2004) on bond markets and Ito, Lyons and Melvin (1998), Evans and Lyons (2002, 2002a, 2005, 2005a), Froot and Ramadorai (2005), Love and Payne (2003), 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).

(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*.

Overcoming this data shortage, we examine a cross-section of traders. In order to identify informed traders we exploit anonymous trader identities on all trades in an electronic limit order interbank currency market. This unique data set enables us to conduct a cross-sectional regression of traders' price impact on six traders' trade characteristics as right-hand side variables. As a side-effect, the consideration of six information indicators in a single regression allows inferences about significance of these indicators when controlling for other relevant variables. 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.

As a first illustration of the characteristics' relation with informed trade we present the following straightforward analysis. We group traders into quintiles ordered by their total transaction volume over the whole sample and calculate the correlation of these groups' order flows with subsequent 10 minute exchange rate returns. Figure 1 shows in an intuitive way: order flow of traders who transact larger volumes has a higher correlation with future price changes than smaller traders, i.e. larger traders provide more information to the market than small traders.

However, we regard this analysis as an illustration only because we aim for overcoming three disadvantages of this method. First, this approach does not really distinguish information from liquidity effects, second, the groups are formed in an arbitrary and non-continuous way and, third, the regression does not control for further potentially important influences. Therefore, we introduce an approach which may be regarded as an extended price impact analysis in the tradition of Hasbrouck's (1991, 1991a) seminal contributions. The extension serves to incorporate trader characteristics into the analysis of informed trading. In order to relate these trader characteristics to traders' price impacts we form random trader groups and estimate the permanent price impact of these groups' order flows. The estimated impacts can then be related to six characteristics—which are commonly thought of as proxying for private information—of these different trader groups.

Thus, our main contribution is to identify informed traders—i.e. traders whose market orders have a large permanent price impact—in a cross-sectional approach 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, trade at times of wide spreads and trade when the order book is thin. It may be reassuring that these results are well in line with microstructure theory and economic intuition. Nevertheless, we do not know of a study where they have been elicited in a cross-section of many traders.

Beyond the empirical identification of single indicators of informed trade, it is another advantage of the data that the multivariate approach reveals the economic significance of each indicator controlling for five further relevant indicators. Moreover, an extension of our standard approach indicates a non-linear effect of the trade size variable as suggested by the so-called "stealth trading" literature (Chakravarty, 2001). We use the highly disaggregated data to demonstrate the effect of "stealth trading"—i.e. informed traders' preference of medium sized trades—in a rigorous way that was not possible before.

Finally, our analysis sheds light on further 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) It is among the first to test the Bloomfield, O'Hara and Saar (2005) experimental finding—trading time can identify informed traders—with real market data.

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 electronic interbank trading. As this 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 has four sections and a conclusion. 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

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,

[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

2004, for a discussion of "splitting orders"). Chakravarty (2001) and Anand and Chakravarty (2005) empirically investigate 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.

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) (Sias and Starks, 1997, Chakravarty, 2001, Jones and Lipson, 2004, Sias, Starks and Titman, 2006). 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.

A third potentially important variable is *local proximity of a trader to a financial or economic center*. Empirical evidence for equities (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

---

Glosten, 1994, Biais, Hillion and Spatt, 1995, Evans, 2002 or Hollifield, Miller and Sandås, 2004).
[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

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—being also closer to the central bank—to have superior information since they have better access to order flow from financial customers. Hence, price impact and local proximity to financial centers 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) develop this idea from the theoretical literature and experimentally show that informed traders tend to trade early in a trading session to capitalize on their private information before other traders can exploit the 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.

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 points out the importance of 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 partially 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.

---

one large trade. Therefore, average trade size of a trader and his overall size are not necessarily the same.

The last variable of interest is the *outstanding order book volume*. Higher liquidity naturally alleviates the short-run price impact of order flow. Here, however, we analyze exclusively the permanent price impact—reflecting information only—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 Data, cross-sectional approach 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 to serve as a country-wide platform for traders from all over Russia. This is important because before introduction of the UTS electronic currency trading in Russia occurred 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. Due to its comparatively high liquidity the UTS accounts for the largest share of Rus-

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

sia'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 of our sample, trading at the UTS was limited to one hour per day, starting 10.30 Moscow time. Nowadays, electronic trading is extended to RUR/EUR and is prolonged to three hours per day. Figure 2 shows the resulting exchange rate over the nine days examined here.

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. This data also allows 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. This is important since traders naturally differ in their size and sophistica-

[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.

tion. In Figure 3 we document the distribution of trader size by plotting all 723 traders on the horizontal axis (sorted by size) and their cumulative market share for total trading volume on the vertical axis. As can be seen there is a strong concentration of trading volume so that e.g. the hundred largest traders account for 50% of total trading volume.

In addition, we are able to determine which of the eight regional exchanges a trader belongs to. In the following, we will refer to traders from Moscow and St. Petersburg as traders from financial centers (FC-traders) and to the traders from the remaining six exchanges as traders from outside financial centers (NFC-traders).[8]

Table 1 shows some descriptive statistics for our trading data from the UTS. Here and throughout the rest of the paper all overnight returns are eliminated. The data set contains 14,109 market orders with a mean transaction size of about 50,000 USD, which is much lower than average trade sizes in the major markets, such as the EUR/USD market, where mean trade sizes typically exceed one million USD. The percentage spread averages 0.0071, which is 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 of 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 orders. The latter is different 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 JPY/USD trading on EBS by Ito and Hashimoto (2004).

[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.

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 foreign exchange 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 Describing the cross-section of traders

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 liquidity effects. We argue that among the menu of available empirical approaches to conduct the cross-sectional analysis the randomization chosen serves its purpose best.

We have to calculate the price impact for every trader in our sample to relate this individual price impact to characteristics of this trader. While these individual price impacts might be obtained by simple regression approaches, e.g. unbalanced panel regressions of midquote returns on individual order flow, these procedures have a severe shortcoming: they do not separate permanent—and thus information based—price impacts from transitory impacts due to inventory or liquidity effects. A well known empirical setup to overcome this problem is the SVAR approach of Hasbrouck (1991) which serves to measure the permanent impact of order flow on asset returns. We will thus use this approach here to obtain a measure for information based price impacts. Another benefit is of this method is to ensure comparability of our results with other papers attempting to single out the information effect of order flow on asset price dynamics (e.g. Brandt and Kavajecz (2004) for bonds, Hasbrouck (1991) for equities and Payne (2003) for foreign exchange).

The measurement of permanent price impacts is thus crucial. However, estimating SVARs for traders individually is not feasible empirically, since we have a lot of small traders with very few trades that are far apart in time. It would be possible to drop these traders from the analysis and estimate the SVARs on traders with sufficient observations only, but this would results in a loss of sample information due to a biased sample of large traders only. Therefore, we rely on forming groups of traders for which the Hasbrouck-SVAR can be estimated and for which we measure average group characteristics such as the average trader size, the share of traders that belong to a financial center and so forth.

In order to disentangle this grouping approach from our subjective influence we rely on a randomization of trader groups that takes independent draws from the trader population and admits a test of the null of identical permanent price impacts across different types of traders. We next describe this randomization procedure in detail.

For each run of our randomization procedure we randomly assign each trader from the population of all 722 traders 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, at each run 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 (Trade Size) of the group (i.e. mean transaction volume per trade and trader),

2. the average market share of traders in this group (Trader Size), i.e. the average of the market share (in terms of trading volume) of each trader belonging to this group,

3. the share of traders in the group that are located in a financial center (Fin Ctr),

4. the volume-weighted minute of the trading session (Time), the group trades at (measured in event time),

5. the average volume-weighted bid-ask spread (Spread) 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 (Book Vol) just prior to the trades of the group (measured in event time).[9]

Conceptually following Hasbrouck (1991a, b) we run a structural VAR with spot midquote returns, order flow from traders in the randomly formed group one (x^1) and order flow from traders in group two (x^2) as independent variables of the following form:

$$Ay_{t+1} = \Gamma(L)y_t + B\upsilon_{t+1} \quad \text{with } \mathrm{Var}[\upsilon_{t+1}] = I_3 \tag{1}$$

where

$$A = \begin{pmatrix} 1 & -\alpha_1 & -\alpha_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, y_t = \begin{pmatrix} r_t \\ x_t^1 \\ x_t^2 \end{pmatrix} \text{ and } B = \begin{pmatrix} \beta_r & 0 & 0 \\ 0 & \beta_1 & 0 \\ 0 & 0 & \beta_2 \end{pmatrix} \tag{2}$$

and r denotes midquote spot returns. $\Gamma(L)$ 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 $\Gamma(L)$. Compared to the original setup in Hasbrouck (1991, 1991a), we have added a second order flow variable which makes the system overidentified with one degree of freedom. However, this setup most naturally corresponds to the original parameterization and we want to be able to compare our results to the earlier literature. Furthermore, the validity of this restriction can be tested and we do this as we proceed. Since we are primarily interested in the total price impact per USD

[9] Volume-weighting is done with respect to the group under consideration, i.e. each observation of a group *i* is weighted by that observation's volume with respect to the total trading volume of that group and *not* with respect to total trading volume of all traders.

we compute the analytical long run price impact of order flow, Ξ_∞, by inverting the VAR operator[10] (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.

Having presented the randomization procedure, two main advantages over competing approaches might be emphasized: it directly allows for a multivariate assessment of the relative importance of the six characteristics and it enables a straightforward application of the Hasbrouck-SVAR. In our view, the latter argument is highly important since this will ensure comparability with earlier results from the literature and since it is the established way to calculate permanent price impacts. In order to highlight these two advantages further, consider an often chosen alternative to our randomization procedure: group traders (e.g. into quintiles) along one dimension (trader size, say) and estimate the permanent price impact for each group. Repeating this for each desired trader characteristic also yields a cross-section of price impacts, but there is no direct way to investigate the relative importance of the different trader characteristics for their influence on the permanent price impacts. Furthermore, it is not clear how to estimate the SVAR on these five groups since the VAR system will be even more heavily overidentified than in the bivariate case in equation (2) unless one is willing to make some more or less arbitrary identifying assumptions. While two- or three-way sorting approaches circumvent some of the problems when attempting to measure the relative impor-

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

tance of different trader characteristics, these sorts would further increase the identification problem.[11]

Our approach circumvents all these problems and yields a simulated distribution of price impacts with minimal overidentifying restrictions. It also allows the use of straightforward regression approaches to measure the influence of the six trader characteristics on permanent price impacts since the procedure yields a cross-section of price impacts and corresponding trader characteristics. Finally, by randomizing over the composition of groups and the size of the two groups we can generate substantial variety in the trader characteristics of single groups. By pure chance, there will be some groups consisting of large traders who use relatively small trade sizes or traders from a financial center that have a relatively small size compared to the market. This dispersion in the underlying characteristics then allows for a reliable assessment of the relative importance of different trader characteristics for the cross-section of price impacts.

Descriptive statistics for these price impacts and group characteristics can be found in Appendix 1. Mean trade size for example ranges from 28,000 to 75,000 USD, the share of traders from a financial center covers groups from 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 any conventional level of significance. As a robustness check, we also allow for correlation between the order flows from the two groups which just-identifies the system in (2). However, the following results are nearly unchanged.

[11] One could also think of using (unbalanced) panel regression methods to allow for trader heterogeneity. However, it is not clear how to estimate the cross-section of permanent price impacts with a reasonable degree of accuracy and how to make the individual effects dependent on trader characteristics without using simulation techniques anyway.

Given these preliminary steps—introducing the data, approach and descriptive statistics—we are now able to relate price impacts to the six group characteristics. This allows us to draw inference about factors that influence the degree of informed trading.

4 Results

This section presents results in an order of increasing complexity and specificity. We start with simple bivariate relations before the relevant multivariate results are presented (4.1). Among the latter we use a weighted least squares regression analysis as benchmark (4.2) whereas robustness tests are discussed in Section 5. As non-linearity seems to be effective we use kernel regressions to examine the role of trade size for informed trading in more detail (4.3). Overall, findings for our market are nicely in line with the literature.

4.1 Bivariate relations

First, we look at simple bivariate correlations of price impacts and the six characteristics obtained for each trading group in order to present a full picture of our data set and to produce findings which may be compared to other bivariate relations found before.

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 our trading time variable, 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, consistent with 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—but does not contradict—earlier results obtained in a time-series setting (see inter alia Hasbrouck and Seppi, 2001, Brandt and Kavajecz, 2004, or Payne, 2003).[12] It has to be kept in mind that we calculate bivariate relations in a cross section of trader groups which is different from the earlier results based on time series analyses. 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 period, whereas time-series studies look at price impacts at certain times of the sample. Thus, 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 received 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

This sub-section tackles the question of which traders' characteristics move prices by using weighted least-squares (WLS) regressions. As it turns out, all six trader characteristics influence the price impact in a way consistent with microstructure theory and/or economic intuition. Most notably, the liquidity proxy, outstanding order book volume, has a negative impact on the informativeness of a trader group's trades in a multivariate context as predicted by Admati and Pfleiderer (1998).

[12] 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).

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.[13] 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 and R^2s show that about 42% of the variation is explained by these six items. Therefore, larger traders from financial centers who trade early, when the spread is high and when order book volume is low have the highest price impact. Interestingly, controlling for trader size, the share of traders from financial centers and trading time indeed leads to a negative sign for outstanding order book volume and a positive sign for the bid-ask spread. This might best be explained by correlated trading strategies by informed traders. Trading of informed traders leads to a reduction in market liquidity since uninformed liquidity traders avoid to trade in times of high information risk. Consequently, a higher spread and a lower book volume signals the presence of informed traders as hypothesized in theoretical work (e.g. Admati and Pfleiderer, 1988).

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

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

alone (Bjønnes and Rime, 2005). However, in light of Chakravarty's findings on stealth trading for stock markets (Chakravarty, 2001) we add a squared trade size variable to the regression and the results are shown in column (2). The squared trade size variable enters negatively, implying an inverted 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 foreign exchange markets, too. Furthermore, in this specification, the three information variables "Trader Size", "Fin Ctr" and "Time" 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, Trader Size and share of traders from financial centers (Fin Ctr). 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 (see Table 3) translates into a "base 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 foreign exchange markets (Payne, 2003, Bjønnes and Rime, 2005).

As can be inferred from Table 4, both Trader Size and the share of traders from a financial center (Fin Ctr) 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 Stealth trading and combined effects of trader characteristics

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

As noted earlier, 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, confirming the so-called "stealth trading" hypothesis of Barclay and Warner (1993), which holds that informed traders choose medium-sized orders. This permits them to avoid giving away their information too easily, which would occur if they were to trade very large orders. It also permits them 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 foreign exchange 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.[14] 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

$$PI_i = g(X_i) + \xi_i \tag{6}$$

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 $E[PI_i \mid X_i]$ 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.

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

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

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

where h is a bandwidth parameter, $K(\cdot)$ is a Gaussian product kernel, 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, so $h_*\approx 0.32$. 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.[15]

We now use this approach to take a look at the interaction of trader size, share of traders from financial centers, 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 4, Panel A, shows a surface plot of price impacts as a function of Trade Size (horizontal axis) and Fin Ctr—the share of traders from financial centers—(vertical axis), i.e. we estimate (7), set all variables except Trade Size and Fin Ctr at their unconditional mean

[15] 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.

and plot fitted price impacts for different combinations of Trade Size and Fin Ctr. In Figure 4, 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 for small or large trade sizes, which confirms the stealth trading hypothesis as discussed above for foreign exchange 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 traders from financial centers (Fin Ctr) and 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. In our case, information is the commodity to be produced, whereas Trader Size and Fin Ctr are the production factors: Trader Size and Fin Ctr "produce" information with marginally declining returns and are substitutes.

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, an adjustment of order flow for autocorrelation in individual order flow and a different estimation approachs. 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 often from financial centers, and both should trade 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 share of traders from financial centers (Fin Ctr) is about 0.30, the correlation of outstanding order book volume and share of traders from financial centers (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, 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.[16]

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 interesting 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

[16] A commonly used rule of thumb is that VIFs should be smaller than 10 und condition numbers should not exceed 20. Our values are far below these numbers.

for up-to second order autocorrelation. 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.[17] 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.[18]

Finally, as a last robustness test, we change the *estimation procedure*. Instead of repeatedly estimating the long-run price impact of a random trader group according to equations (1) to (3), we estimate the following forecasting equation for each of the two trader groups in each of the 25,000 runs of the randomization procedure in event time

$$r_{t+1\to t+10} = \beta_0 + \beta_1 x_t^i + \varepsilon_{t+1\to t+10}\,. \qquad (10)$$

where $r_{t+1\to t+10}$ is the return over the next ten minutes and x_t^i is here the order flow of trader group i at time t. Therefore, this test uses tick-by-tick data and estimates the forecasting power

[17] We do this adjustment for each trader who has at least ten trades on a given day.
[18] Results are available from the authors upon request.

of different trader groups' order flow to forecast return. The randomization procedure then gives us 50,000 estimates of β_1 along with the corresponding t-statistics and R^2s.

We present results for regressing the t-statistic of β_1 on the standardized six trader group characteristics in Appendix 3.[19] As the results in column (1) show, all six trader characteristics have the same effect of a group's forecasting power as discussed above, e.g. larger traders, traders from financial centers etc. convey more information. For example, a one standard deviation shock in trader size raises the t-statistic by more than two points. Column (2) excludes the bottom 5% and top 5% most extreme groups in terms of their estimated β_1 to control for outliers. As can be inferred, outliers do not seem to drive our results.

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

[19] We choose the t-statistic to account for the uncertainty in the forecasting power of a group's forecasting power. However, using the estimate of β_1 does not change our result.

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, simultaneously 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. These findings are robust to several modifications, such as splitting the sample, using a refined order flow measure or when explaining future price changes.

As a focus of our research, we examine "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-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 testing the findings for different data or institutional settings and 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.

Anand, Amber and Sugato Chakravarty (2005), Stealth Trading in Options Markets, *Journal of Financial and Quantitative Analysis*, forthcoming.

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.

Bozcuk, Aslihan and Ameziane Lasfer (2005), The Information Content of Institutional Trades on the London Stock Exchange, *Journal of Financial and Quantitative Analysis*, 40, 612-644.

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 (2007), *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.

Hautsch, Nikolaus and Dieter Hess (2004), Bayesian Learning in Financial Markets: Testing for the Relevance of Information Precision in Price Discovery, *Journal of Financial and Quantitative Analysis*, forthcoming.

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.

Love, Ryan and Richard Payne (2003), Macroeconomic News, Order Flows, and Exchange Rates, *Journal of Financial and Quantitative Analysis*, forthcoming.

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 (2006), Changes in Institutional Ownership and Stock Returns: Assessment and Methodology, *Journal of Business*, 79, 2869-2910.

Figure 1. Correlation of order flow and future price changes by traders sorted on size

This figure shows the correlation of market order flow (measured in dollars) with future 10-minutes imdquote changes for different groups of traders. Traders are sorted into quintiles according to the size of their total trading volume. Q1 comprises the largest traders whereas Q5 is made up by the smallest traders that account for 20% of total trading volume, respectively.

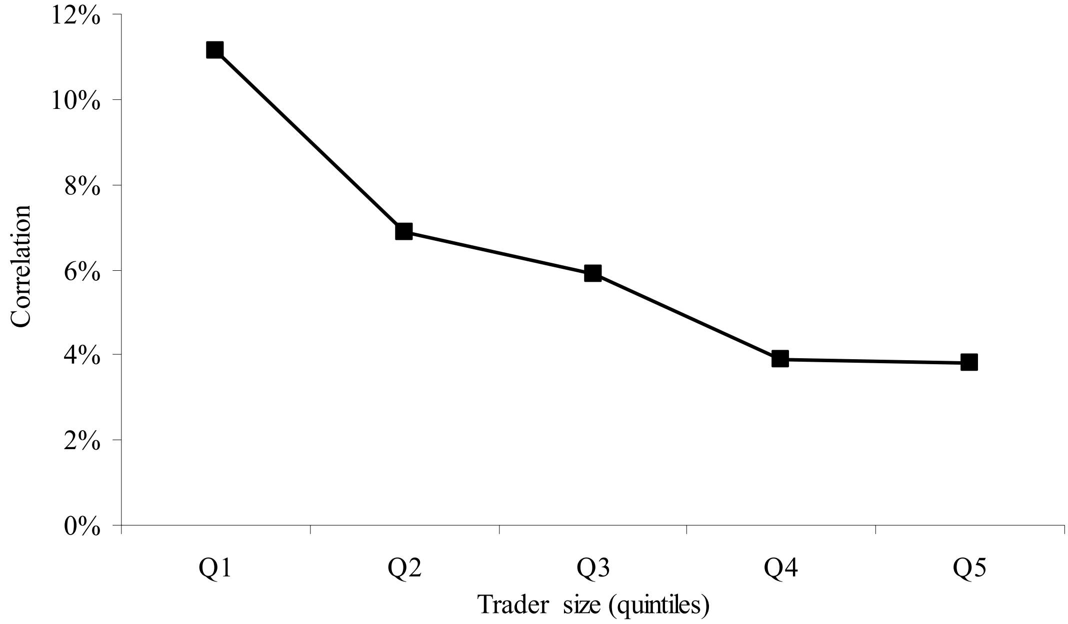


Figure 2. RUR/USD spot exchange rate

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).

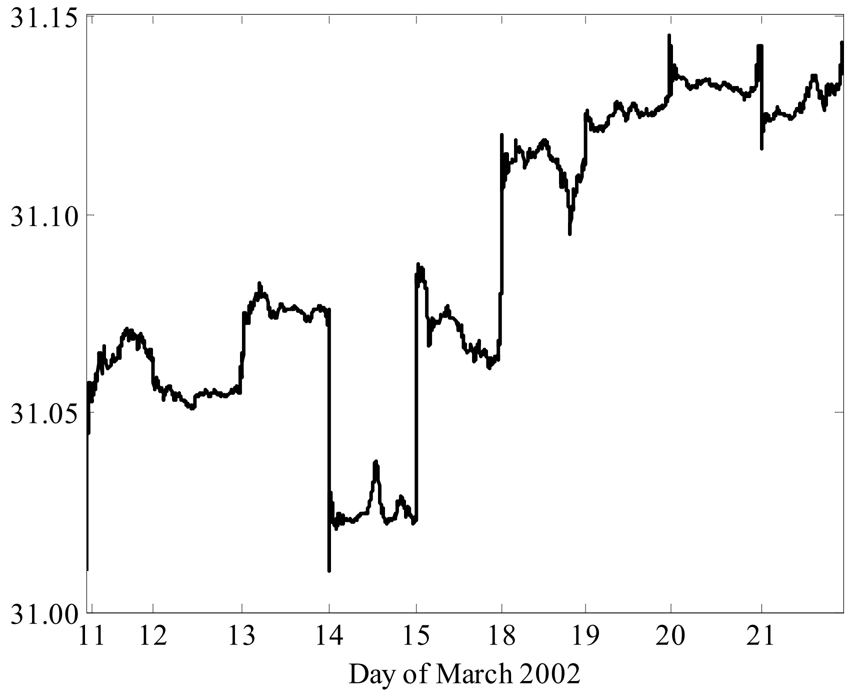


Table 1. Descriptive statistics

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 auto-correlation 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 3. Distribution of trading volume

This figure shows the cumulative share of total trading volume for all 723 traders in the sample. Traders (x-axis) are sorted by size in descending order.

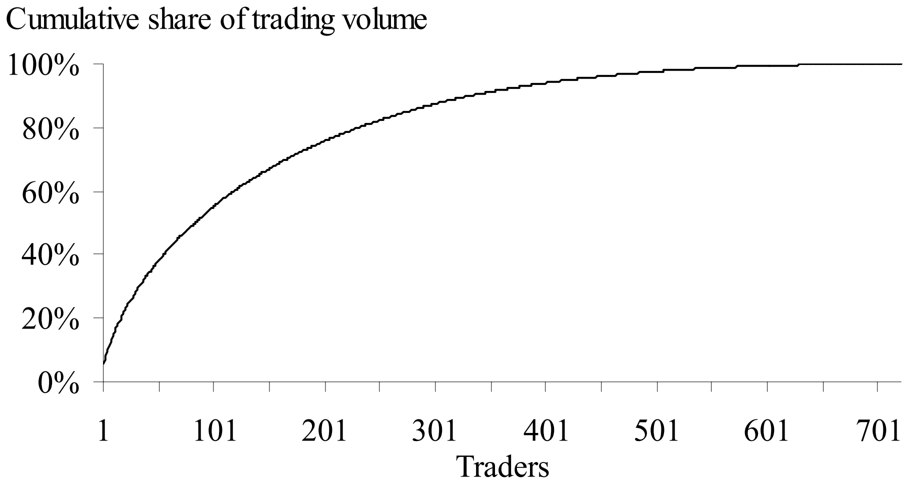


Table 2. Correlation of price impacts and trader group characteristics

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

	Price Impact	Trade Size	Trader Size	Fin Ctr	Time	Spread
Trade Size	0.22					
Trader Size	0.26	0.41				
Fin Ctr	0.28	0.17	0.30			
Time	-0.24	-0.12	-0.08	-0.37		
Spread	0.59	0.34	0.16	0.07	-0.19	
Book Vol	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 ($\times 10^{-5}$), (average) trader size calculated as the average market share of traders in a respective group (in %), share of traders from a financial center (Fin Ctr), trading time, bid-ask spread, and outstanding order book volume. 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)
Trade Size	-0.015 (-0.90)	0.026 (1.47)	0.135 (5.56)			0.042 (2.22)
$(\text{Trade Size})^2$		-1.230 (-3.28)	-1.748 (-13.24)			-1.354 (-3.21)
Trader Size	0.064 (4.15)	0.051 (3.46)	0.046 (2.34)	0.086 (4.83)	0.084 (4.75)	0.053 (3.83)
Fin Ctr	0.133 (5.67)	0.129 (6.33)	0.137 (4.97)	0.154 (4.07)	0.168 (4.53)	0.128 (6.35)
Time	-0.003 (10.66)	-0.002 (-2.86)	-0.002 (-2.25)	-0.002 (-2.69)		-0.002 (-2.62)
Spread	0.003 (10.66)	0.003 (13.36)				0.003 (9.58)
Book Vol	-0.021 (-3.89)	-0.017 (-4.48)				-0.019 (-4.99)
Const.	0.026 (50.41)	0.026 (49.92)	0.026 (31.21)	0.026 (26.05)		0.026 (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 (0.00)	55.86 (0.00)	43.38 (0.00)	28.10 (0.00)	25.64 (0.00)	38.46 (0.00)
obs	50,000	50,000	50,000	50,000	50,000	40,000

Table 4. Economic significance

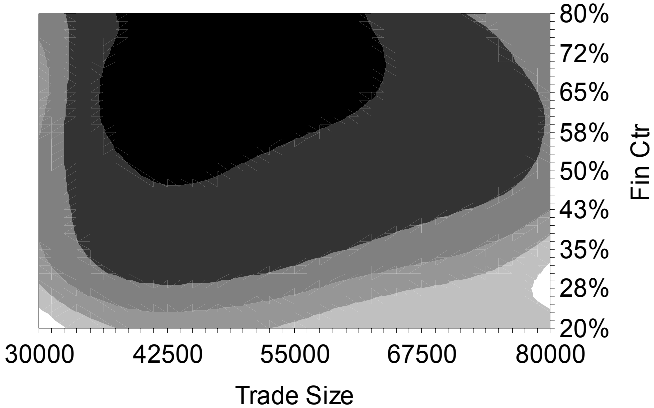
This table shows the effect of a one standard deviation increase in one of the six trader group characteristics on the percentage deviation of price impacts from their conditional mean while holding all other five characteristics fixed. The five columns show the effect on price impacts according to the 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
Trade Size	-2.22%	-11.51%	-1.53%			-10.72%
Trader Size	7.07%	5.55%	4.93%	9.36%	9.23%	5.83%
Fin Ctr	23.54%	22.57%	23.61%	26.71%	29.11%	22.46%
Time	-12.39%	-8.14%	-5.57%	-7.38%		-7.20%
Spread	32.11%	32.46%				27.97%
Book Vol	-21.83%	-17.23%				-19.24%

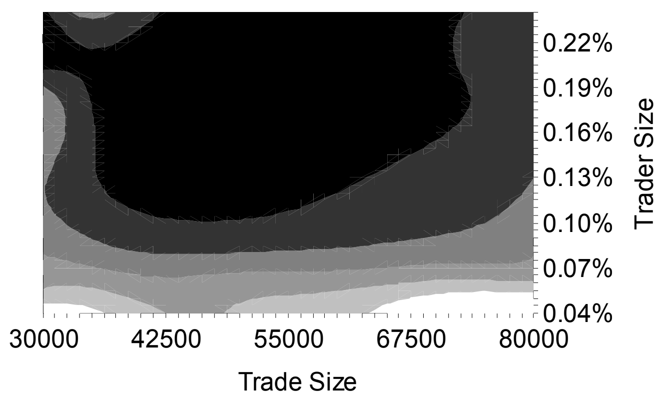
Figure 4. Kernel regression: price impacts, trade size, FC traders and trader size

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 and the share of traders from a financial center (Fin Ctr). Panel (B) shows price impacts depending on trade size and trader size. Panel (C) shows price impacts for different combinations of the share of traders from a financial center (Fin Ctr) and trader size.

Panel A: Trade Size and Fin Ctr



Panel B: Trade Size and Trader Size



Panel C: Trader Size and Fin Ctr

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 ($\times 10^{-5}$), (average) trader size calculated as the average market share of traders in a respective group (in %), share of traders from a financial center (Fin Ctr), trading time, bid-ask spread, and outstanding order book volume. q(·) denotes the quantile for the value in brackets.

	Price Impact	Trade Size	Trader Size	Fin Ctr	Time	Spread	Book Vol
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 ($\times 10^{-5}$), (average) trader size calculated as the average market share of traders in a respective group (in %), share of traders from a financial center (Fin Ctr), trading time, bid-ask spread, and outstanding order book volume. 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
Trade Size	1.35
Trader Size	1.30
Fin Ctr	1.41
Time	1.97
Spread	1.27
Book Vol	2.17
Condition number	2.88

Appendix 3. Cross-sectional results from forecasting regressions

This table shows results for regressing the t-statistic for the coefficient β_1 from the forecasting equation

$$r_{t+1\rightarrow t+10} = \beta_0 + \beta_1 x_t^i + \varepsilon_{t+1\rightarrow t+10}$$

on six trader characteristics. All explanatory variables are standardized. The second column excludes the bottom and top 5% groups with the most extreme estimates of β_1.

	(1)	(2)
Trade Size^2	-1.78	-1.77
	(-43.60)	(-38.55)
Trader Size	2.41	3.39
	(47.95)	(42.45)
Fin Ctr	1.99	1.54
	(3.63)	(4.84)
Time	-0.95	-0.91
	(-3.14)	(-3.02)
Spread	2.51	2.51
	(7.30)	(6.61)
Book Vol.	-1.12	-1.09
	(-2.15)	(-1.99)
Const.	4.03	4.03
	(24.12)	(27.26)
adj. R^2	0.13	0.13
AIC	4.49	4.45
obs	50,000	40,000