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 a foreign exchange 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 which—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.

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

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. 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. Although theory has provided guidance as to which kind of traders might be informed, the comprehensive empirical identification of informed traders is still listed among the big open questions in the microstructure literature (Lyons, 2001, Hasbrouck, 2007).

Available evidence of informed trade in foreign exchange consists of either single dealer studies (recently e.g. Bjønnes and Rime, 2005), studies of the aggregate market (e.g. Evans and Lyons, 2002, Payne, 2003) or studies of end-user customer groups (e.g. Lyons, 2001, Evans and Lyons, 2005a), but we do not know of a single study where a large cross-section of traders has been analyzed.[1] Accordingly, there is no cross-sectional evidence on characteristics of informed versus uninformed traders in the foreign exchange market. Among possible characteristics of informed traders, there are three important ones which have not yet been examined in foreign exchange. First, we do not know whether large trade size is a useful proxy for identifying informed traders or whether "stealth trading" is performed as in other markets (e.g. Chakravarty, 2001). Second, we do not know whether the strongly-voiced opinion that large traders are better informed (e.g. Cheung and Chinn, 2001) holds an econometric test. Third, there is no evidence as to whether the experimental revelation of early trading

---

[1] See studies on stock markets, such as Chan and Lakonishok (1995), Madhavan and Cheng (1997) or Lei and Wu (2005) and on bond markets see Massa and Simonov (2003).

time as a characteristic of informed traders applies to foreign exchange (Bloomfield, O'Hara and Saar, 2005).

Beyond these three possible characteristics of informed traders there are three further ones which have been found in earlier studies, i.e. traders' location in a financial center, high bid-ask spreads and high or low book volume. This variety of characteristics nurtures the concern whether all of these characteristics are still relevant in a multivariate approach. Our ambition of characterizing informed traders creates an almost natural research design in this respect, i.e. to relate the traders' permanent price impact—indicating their degree of information—to their characteristics.[2] From a unique data set we exploit anonymous trader identities of all trades in an electronic limit order interbank currency market. These data enable us to conduct a cross-sectional regression of traders' price impact on six traders' trade characteristics as right-hand side variables.[3]

Our empirical approach may thus be regarded as an extended price impact analysis in the tradition of Hasbrouck's (1991, 1991a) seminal contributions. The extension serves to incorporate several relevant trader characteristics in 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 we control each proxy of informed trade by a set of five relevant determinants. The cross-sectional approach reveals that information is conveyed by traders who—simultaneously—trade medium-sized orders, have large

---

[2] For the information contained in order flow see, for example, 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, Hasbrouck (1991, 1991a, 2007), Dunne, Hau and Moore (2004), or Engle and Patton (2004) on stock markets, and Brandt and Kavajecz (2004), Green (2004) on bond markets.

[3] 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).

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. Reassuringly, this result does not depend on our preferred method but also shows up if we apply two variants of a predictive regression approach.

Thus the cross-sectional approach brings new insights to the foreign exchange literature which nevertheless seem to be well in line with microstructure theory and economic intuition. First, there is no other study documenting stealth trading in foreign exchange. Second, the new analysis of trader size obviously profits from trader-related micro data which were not available before. Third, the time of trading in the sense used by Bloomfield, O'Hara and Saar (2005) has not been analyzed before as it needs joint information about informed traders and their trading times. Fourth, we do not know of another empirical (cross-sectional) study where many relevant characteristics have been examined in a single framework, thus allowing us to control for possibly competing influences.

Earlier work on informed trade in foreign exchange has either relied on single dealer studies which do not allow for a comparison with the market, or it has relied on groups of market participants, which does not allow comparison of traders with each other, or it focuses on the time-series dimension. The latter provides related but different information, as we argue below. For example, Payne (2003) is possibly the first to apply the Hasbrouck approach to identify informed trade in the foreign exchange market and then to test whether periods of more informed trade are related to a particular trade size. Owing, however, to the aggregated kind of data available, this analysis based on the average degree of informed *trade* can at best indicate whether any relation found stems from the behavior of informed *traders*.[4] A more

---

[4] Payne (2003, p.322) does not find a relation in his data. This is consistent with—but cannot be identified as—stealth trading because informed traders prefer an average trade size.

convincing analysis of possible stealth trading requires disaggregated data in order to compare the preferred trade size of informed versus uninformed traders.[5]

Owing to the two-tier market structure in foreign exchange, respective studies may differ from those in stock markets. Our data cover the interbank market, as do, for example, Lyons (1995), Evans and Lyons (2002) and Payne (2003). Much of the information in foreign exchange, however, seems to stem from informed customers and is thus first revealed in the customer market, i.e. is exclusively between customer and bank, and then transported by the banks' traders into the interbank market. We do not assume a particular mechanism by which information gets incorporated in prices but implicitly assume that traders' behavior reflects information independent of its origin.[6]

This study is based on a new database, i.e. an anonymous but otherwise complete record of transactions in a modern pure limit order market.[7] 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 those of 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 on 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.

---

[5] See Chakravarty (2001), who shows that there are pronounced differences between investor groups.

[6] The referee correctly points out that the widespread simultaneous trade model used e.g. in Evans and Lyons (2002) has a specific structure in which traders transport customers' information by balancing their inventories at the end of three stylized trading rounds. In contrast, we do not focus on the markets' two-tier structure and how to derive market equilibrium in this setting but we apply a reduced form approach similar to that of the original empirical stock market literature.

[7] Electronic limit order books 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 and Hollifield, Miller and Sandås, 2004).

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). One would therefore expect a positive relation between the 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. This traditional indicator of informed trade, however, 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) and Anand and Chakravarty (2005) empirically investigate the effect of "stealth trading" (Barclay and Warner, 1993) in stock and options markets and find 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 is therefore possibly non-linear.

A second and closely-related variable is *trader size*.[8] 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 actually do believe that large players are more informed (see Cheung and Chinn, 2001). Early evidence on a potentially

---

[8] 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. Average trade size of a trader and his overall size are therefore not necessarily the same.

important role of larger traders was provided by Peiers (1997) in an analysis of the role of single large banks in leading the market.[9] 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*. 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).[10] 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. It has indeed been shown for our market that price impact and local proximity to financial centers are positively correlated (Menkhoff and Schmeling, 2007) but it is an open question whether this also holds in the multivariate approach applied here.

The *time of day* at which a trader places 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. Earlier trading is therefore a proxy for superior information and we expect price impacts to be the higher the earlier a trader places his orders. The time-series study of Payne (2003) has indirectly analyzed this relation as he finds that the average price impact of a trade is lower in the (European) morning of DEM/USD trading when the market is most active. This evidence seems inconsistent with the hypothesized time-of-day effect but Payne (2003) also

---

[9] 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 with small traders (individuals) (Sias and Starks, 1997, Chakravarty, 2001, Jones and Lipson, 2004, Sias, Starks and Titman, 2006).

[10] 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.

finds that trading overall is more informative in the morning hours, indicating that informed traders are very active in the market at that time.

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 compensate partially for adverse selection in currency markets.[11] From these earlier findings, we expect the spread at which traders place their orders to influence the overall price impact of orders positively.

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. One should therefore expect to find a negative relation between the level of liquidity a trader prefers to trade at and his price impact (see Payne, 2003). As also discussed in Admati and Pfleiderer (1988), however, higher liquidity could attract informed traders attempting to trade at low costs. Traders who trade when order book volume is comparatively high might therefore show up with higher price impacts in a

---

[11] In the interbank market covered by Payne (2003) spread serves the conventional role of protecting against better-informed counter-parties. In the customer market covered by Osler, Mende and Menkhoff (2006), however, spread indicates customers' information which is accordingly honored by the banks. As we cover the interbank markets here, we expect that higher spreads are related to more information.

cross-sectional analysis. The relation between the level of liquidity at which a trader prefers to trade and his price impact are therefore *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).[12] 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 in eight regional exchanges which were not linked to each other. Traders from e.g. Moscow could not therefore trade electronically with traders from e.g. Rostov. Owing to its comparatively high liquidity the UTS 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.

Thus, the UTS represents domestic trading but does not represent foreigners who traded roubles in an offshore market in the form of non-deliverable forward contracts. Accordingly, the UTS data cannot cover the total RUR/USD market but they completely cover a relevant segment.[13]

---

[12] Goldberg and Tenorio (1994) also analyze trading at the MICEX. Their data do not, however, come from the modern electronic trading system.
[13] This limitation applies to most foreign exchange studies because there are usually several parallel trading venues, such as (various) electronic markets, voice brokers and direct interbank trading.

Trading takes place in an electronic limit order book called SELT which is very similar to the systems of Reuters and EBS.[14] At the time of our sample, trading at the UTS was limited to one hour per day, starting 10.30 a.m. Moscow time. Nowadays, electronic trading is extended to RUR/EUR and is extended to three hours per day. Figure 1 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 submission of marketable limit orders.[15] The limit order book has strict 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 counter-parties after completion of 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. This is important since traders naturally differ in their size and sophistication. In Figure 2 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

---

[14] SELT was in fact developed in cooperation with REUTERS.
[15] Payne (2003) and Hasbrouck and Saar (2007) also treat marketable limit orders as market orders.

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 to which of the eight regional exchanges a trader belongs. 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).[16]

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 quoted 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 because 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).

Midquote returns, also displayed in Table 1, are mean zero, heavily fat-tailed, and show significant negative first order autocorrelation as documented in earlier microstructure analy-

---

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

sis 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 with currency trading in established markets, which implies that insights gained from trading at the UTS might carry over to trading in other markets and assets as well.[17]

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 our 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 mid-quote 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 owing to inventory or liquidity effects. A well-known empirical set-up 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 that it ensures comparability of our results with other papers attempting to single out the information effect of order flow on asset price dynamics (e.g. Payne (2003) for foreign exchange, Brandt and Kavajecz (2004) for bonds, Hasbrouck (1991) for equities).

---

[17] We are aware of the limited size of this sample but could not get more data. Fortunately, the data show an interesting, plausible and stable structure which lends some credence to general results drawn from them.

The measurement of permanent price impacts is thus crucial. Estimating SVARs for traders individually is not feasible empirically, however, 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 result in a loss of sample information and to a biased sample of large traders only. We therefore 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 groups so constructed consists of no more than 90% of all traders and not less than 10% of all traders. At each run we therefore 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 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 is 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).[18]

Conceptually following Hasbrouck (1991a, b) we run a structural VAR with spot mid-quote 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:

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

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 set-up 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 with 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. This set-up, however, most naturally corresponds to the original parameterization and we want to be able to compare our results with those of 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

---

[18] 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.

total price impact per USD we compute the analytical long run price impact of order flow, Ξ_∞, by inverting the VAR operator[19] (see e.g. Lütkepohl, 2005)

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

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.

The procedure's 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 it is also 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

---

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

when we attempt to measure the relative importance of different trader characteristics, these sorts would further increase the identification problem.[20]

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 with 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 of 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). The following results are, however, nearly unchanged.

---

[20] One could also think of using (unbalanced) panel regression methods to allow for trader heterogeneity. It is not, however, 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 statis-
tics—we are now able to relate price impacts to the six group characteristics. This allows us to
draw inferences 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 (4.1) before the relevant multivariate results are presented.
Among the latter we use a weighted least squares regression analysis as benchmark (4.2),
simple characteristic-based sorts and a predictive regression approach (4.3), and further ro-
bustness 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 (4.4). 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 with 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. Although this finding is consistent with e.g. Easley and O'Hara (1992), it stands in sharp contrast to—but does not contradict—earlier empirical results obtained in a time-series setting (see *inter alia* Hasbrouck and Seppi, 2001, Brandt and Kavajecz, 2004, or Payne, 2003).[21] 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. A clearer picture might therefore be received by the measurement of the 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

---

[21] 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).

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).

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). We therefore 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, owing 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.[22] Column (1) shows regression results when all six items are used 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. Larger traders from financial centers who trade early, when the spread is high and when order book volume is low, therefore have the highest price impact. Interestingly, controlling for trader size, the share of traders from financial centers and trading time indeed lead 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 trading in times of high information risk. Consequently, a higher

---

[22] 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.

spread and a lower book volume signal 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 alone (Bjønnes and Rime, 2005). In the light of Chakravarty's findings on stealth trading for stock markets (Chakravarty, 2001), however, 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.[23] 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. 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 (90th) percentile of the empirical price impact distribution. This serves to eliminate the most extreme price impacts which might have been obtained by our random sampling of very extreme groups. It is obvious from the results in this restricted sample that our conclusions are not driven by outliers. Finally, we include two additional control variables in column (7): trading volume and the number of trades over the minute just prior to a trade. While these

---

[23] Our procedure thus uses a continuous measure of trade size and we do not have to sort trades into more or less arbitrary size categories.

variables are not directly grounded in microstructure theory, the general pace of market activity as measured by these variables may also affect the price impact of a trader's trades (see e.g. Payne, 2003). The results show that these two controls do not significantly affect the results from earlier specifications, which is probably because we already incorporate spreads and outstanding order book volume.

Apart from being statistically significant, the results are also of economic relevance. From 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 Characteristic-based sorts and predictive regression analysis

In this section we apply alternative methods to elicit the information content of different trader groups' order flow. These methods are based on analysis of non-modified midquote returns. Qualitative findings confirm results from the WLS regression approach of Section 4.2.

First, we employ a simple univariate disaggregation of traders into groups based on one of the six trader characteristics. Specifically, we compute the value of a characteristic for each of our 722 traders, providing a ranking of all traders for each of the six characteristics. Next we compute the quartile cut-offs for each of the characteristics and group traders into one of the four quartiles of the empirical distribution of trader characteristics. For example, we compute the trader size (total trading volume) for each of the 722 traders, sort traders in ascending order and allocate them to four groups, where the first group comprises the smallest traders that jointly account for 25% of the total trading volume at our market. The second group is made up of the larger traders who also account for 25% of the total trading volume, and so on. The procedure for the remaining trader characteristics is analogous except for the financial center (Fin Ctr) characteristic. Here, we simply allocate traders to two groups based on their affiliation with a financial center or not.

Based on this sorting procedure, Figure 3 shows correlation coefficients of a trader group with midquote returns for each of the six trader characteristics.[24] This simple procedure provides similar main findings to those of the randomization procedure employed in the previous section. Order flow of traders who use very large and small transactions on average has a lower correlation with midquote changes (consistent with stealth trading), larger traders have a higher correlation, etc. Owing, however, to the univariate nature of this sorting, the figure is silent about the relative magnitude of effects and ignores correlations between the six

---

[24] We opt for simple correlation coefficients here to employ a different measure of the information content of order flow.

trader characteristics. We therefore employ two additional robustness tests in the remainder of this subsection that also account for these effects.

Second, we follow the earlier randomization procedure of Section 3.2 but change our measure of price impact. Instead of repeatedly estimating the long run price impact of a random trader group according to equations (1) to (3), we estimate the following predictive equation for each of the two trader groups in each of the 25,000 runs of the randomization procedure in event time

$$r_k^{t+1\rightarrow t+10} = \beta_0 + \beta_{1,k}x_k^i + \varepsilon_k^{t+1\rightarrow t+10}. \tag{4}$$

where $r^{t+1\rightarrow t+10}$ is the return over the next ten minutes and x_k^i is here the order flow of trader group i at trade k. This test therefore uses tick-by-tick data and estimates the forecasting power of different trader groups' order flow to forecast return. The randomization procedure then gives 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 Table 5.[25] 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. Results in column (2) confirm our finding above, namely that the average size of a trader's trades itself is not a good proxy for information as the linear trade size term does not really add explanatory power. Column (3) 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 results. Finally, in column (4) we again include volume and number of trades as additional controls.

---

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

These variables are not, however, significant and do not affect the results regarding the other six explanatory variables.

Third, we also estimate the predictive regression used above but without focusing on trader groups

$$r_k^{t+1 \rightarrow t+10} = \beta_0 + \beta_{1,k} x_k + \varepsilon_k^{t+1 \rightarrow t+10}. \tag{5}$$

on the trade-by-trade data set without aggregating to the one-minute frequency and without the randomization procedure. Here, x_k denotes market order flow which we have standardized to have zero mean and unit variance for ease of interpretation. This regression essentially measures whether order flow at trade k (here k indexes trades in event time) forecasts mid-quote returns over the next ten minutes ($t+1 \rightarrow t+10$). In the specification above, we allow the predictive coefficient $\beta_{1,k}$ to vary over time. Specifically, we estimate several versions of (10) where the coefficient $\beta_{1,k}$ is modelled as a variant of the following specification:

$$
\begin{aligned}
\beta_{1,k} = \gamma_0 &+ \gamma_1 \text{Trade size}_k + \gamma_2 \text{Trade size}_k^2 + \gamma_3 \text{Trader size}_k + \\
&\gamma_4 \text{Fin Ctr}_k + \gamma_5 \text{Time}_k + \gamma_6 \text{Spread}_k + \gamma_7 \text{Book vol} + z_k \delta
\end{aligned} \tag{6}
$$

In this specification, z_k contains further explanatory variables used as controls for the general pace of market activity, i.e. here trading volume and the number of trades over the minute just prior to the trade. Results for different specifications are shown in Table 6.

Specification (1) of Table 6 shows the predictive content of a simple regression of future midquote changes on market order flow. The coefficient on current order flow is highly significant but the adjusted R^2 is low, as can be expected since we are working in event time. Specification (2) also controls for the trading volume and total number of trades over the preceding 20 seconds of trade to net out effects from the general pace of the market.[26]

Specifications (3) to (5) include the cross-sectional determinants in equation (6). Here, a characteristic at event k depends on the dealer completing this trade. For example, if a specific

---

[26] The choice of the 20 seconds interval is arbitrary. Use of intervals of 30 or 60 seconds does not, however, affect our results.

trader trades at event k, the six trader characteristics take on values of that specific trader for this specific trade. We also standardize each of the characteristics (except for the Fin Ctr variable) cross-sectionally in the trader dimension to have zero mean and unit variance for ease of interpretation. The γ coefficients therefore directly show how much the predictive coefficient changes when a characteristic of a trader changes to a one standard deviation above the average trader characteristic.

The results for specifications (3) to (5) show that the predictive content of order flow, measured via β_1, varies significantly with trader characteristics. Corroborating the earlier evidence, order flow seems to be more informative, when a trader uses medium-sized trades, when the trader is larger and from a financial center, when the trader trades earlier in the session and at higher spreads and lower book volume. Similar to the results in Tables 3 and 5, specification (3) shows that linear trade size does not seem to be a powerful determinant of the informativeness of a trader's trades.[27] These findings hold when either the first 15 minutes in specification (6) or only the remaining time in specification (7) are considered.

4.4 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 do 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

---

[27] We have also run regressions underlying Tables 3 and 5 with trading volume and the number of trades prior to a given trader's trade as additional variables. Since, however, these controls are less directly interpretable in the cross-sectional setting underlying our earlier analysis, and since results for our six indicators of informed trade do not change qualitatively, we do not report the extended regressions here. They are, of course, available upon request.

relation of this kind in the empirical literature may be trade size. Empirically, the relation be-
tween 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 initia-
tor of the trades. Medium-sized trades of institutional investors have the highest price im-
pacts, 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 suggest that traders who tend to
place medium-sized trades have the highest average price impact for all their trades on a dol-
lar-per-dollar basis. This section considers a more general set-up, i.e. nonparametric kernel
regressions, to shed light on the influence of trade size in combination with other characteris-
tics on the price impact of a group.[28] This is because Chakravarty (2001) not only finds me-
dium-size trades to be most informative but rather trades of medium size that originate from
institutional investors. A close analogy 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

$$PI_i = g(X_i) + \xi_i \tag{7}$$

---

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

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.

Estimation of the fitted price impact dependent on a particular value of the explanatory

variables x_0 is done the standard way by calculating

$$\hat{g}(x_0; x) = \frac{\sum\limits_{i=1}^{n} PI_i K\left(\dfrac{x_0 - x_{i\bullet}^T}{h}\right)}{\sum\limits_{i=1}^{n} K\left(\dfrac{x_0 - x_{i\bullet}^T}{h}\right)} \tag{8}$$

where h is a bandwidth parameter, $K(\cdot)$ is a Gaussian product kernel, i=1,…,50,000 denotes

observations and equation (8) 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

$$h_* = \left(\frac{4}{m+2}\right)^{\frac{1}{m+4}} n^{-\frac{1}{m+4}} \tag{9}$$

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.[29]

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

---

[29] 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.

according to equation (8), 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 and plot fitted price impacts for different combinations of Trade Size and Fin Ctr. In Figure 3, darker areas mark higher price impacts. As can be seen, price impacts generally increase when more traders from a financial center are added 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 we look 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 with 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 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 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, and 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). As the results in Table 3 show, however, using only a subset of our variables leads to similar results as in the full specification. In addition, owing 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.[30]

As a second sensitivity 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 from 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

---

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

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 adjustment of individual traders' order flow 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,

$$x_k^{adj,i,d} = x_k^{i,d} - \hat{\alpha}_1^{i,d} x_{k-1}^{i,d} - \hat{\alpha}_2^{i,d} x_{k-2}^{i,d} \tag{10}$$

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.[31] 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. All results reported in the preceding sections, however, are virtually unchanged when we use adjusted order flow instead of the usual order flow measure, so we do not report the results.[32]

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.

---

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

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 can still be increased by demonstration of the information aggregation process in more detail. One important step in this direction seems to be comprehensive identification of those traders who bring information into the market. This obviously requires trading data being linked to trader identities. In this respect, our data go beyond available material in foreign exchange—according to the best of our knowledge—and thus allow for analyses that have not been performed before.

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 contribution to the foreign exchange literature is the cross-sectional analysis of informed traders which covers for the first time three *trader's* trade characteristics, i.e. trade size, trader size and time of trade. These characteristics that theory has interpreted as indicators of information are examined simultaneously in a regression with six relevant characteristics overall. The multivariate approach shows 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 trade early in the trading session, and 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 explaining future price changes.

Owing to its prominence in the literature, 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 does not necessarily bring about the same results compared with other approaches. If one relates trade size to price impact (the most conventional approach), there is no (linear) relation in our limit order market. If one relates trade size to traders, how-

ever, 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.

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.

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.

Chan, Louis K.C. and Josef Lakonishok (1995), The Behavior of Stock Prices Around Institutional Trades, *Journal of Finance*, 50, 1147-1174.

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.

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.

Engle, Robert F. and Andrew J. Patton (2004), Impacts of Trades in an Error-Correction Model of Quote Prices, *Journal of Financial Markets*, 7, 1-25.

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 (2007), 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 (2004), 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.

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

Lei, Qin and Guojun Wu (2005), Time-Varying Informed and Uninformed Trading Activities, *Journal of Financial Markets*, 8, 153-181.

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. (1995), Tests of Microstructural Hypotheses in the Foreign Exchange Market, *Journal of Financial Economics*, 39, 321-351.

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.

Madhavan, Ananth and Minder Cheng (1997), In Search of Liquidity: Block Trades in the Upstairs and Downstairs Markets, *Review of Financial Studies*, 10, 175-203.

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.

Massa, Massimo and Andrei Simonov (2003), Reputation and Interdealer Trading: A Microstructure Analysis of the Treasury Bond Market, *Journal of Financial Markets*, 6, 99-141.

Menkhoff, Lukas and Maik Schmeling (2007), Local Information in Foreign Exchange Markets, *Journal of International Money and Finance*, forthcoming.

Osler, Carol, Alexander Mende and Lukas Menkhoff (2006), Price Discovery in Currency Markets, EFA 2006 Zurich Meetings Discussion Paper.

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. 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).
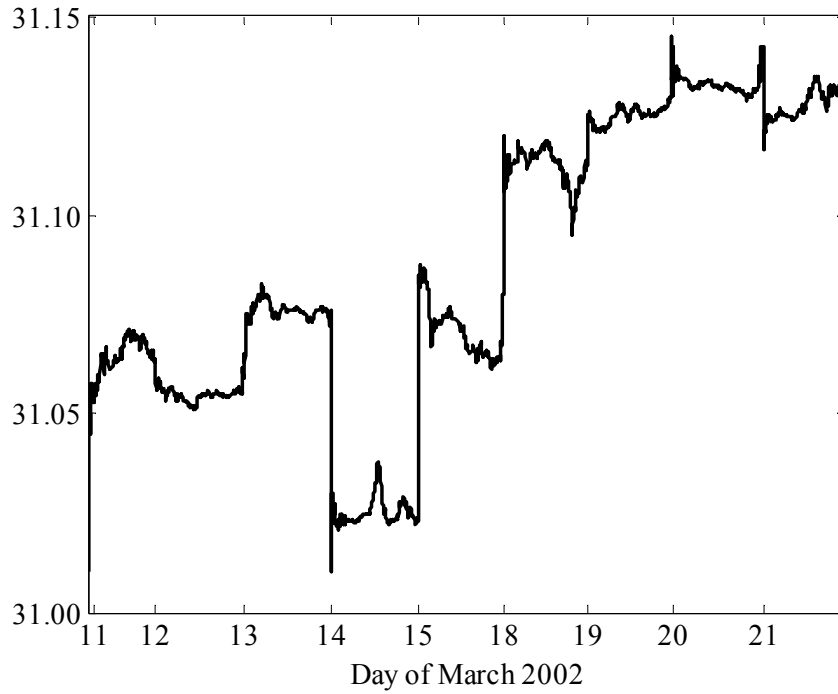

Day of March 2002

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 two columns show the effective and quoted percentage spreads.

	mean ($\times 10^3$)	st. dev. ($\times 10^3$)	skewness	Kurtosis	ρ_{-1}	OBV	OBC	trade size	no. trades	effective pct. spread	quoted pct. spread
All	0. 002	0.301	-0.0802	18.70	-0.0961	17.61	165.96	49,396	14,109	0.0069	0.0071
5	-0. 001	0.276	-0.4299	24.65	-0.1318	19.81	145.35	55,795	3,140	0.0114	0.0115
10	0. 001	0.294	-0.0955	18.80	-0.1070	22.86	204.63	52,236	2,404	0.0043	0.0045
15	0. 013	0.289	0.3732	16.32	-0.1361	19.18	191.60	49,009	1,907	0.0042	0.0043
20	-0. 003	0.290	-0.3718	18.82	-0.0600	18.39	187.51	47,362	1,242	0.0047	0.0049
25	-0. 009	0.299	-0.3256	19.03	-0.0447	15.70	184.41	46,821	1,024	0.0048	0.0049
30	-0. 011	0.308	-0.0777	17.02	-0.0132	15.86	173.98	39,200	832	0.0044	0.0046
35	0. 004	0. 321	-0.0760	15.96	-0.1488	15.00	171.48	44,903	585	0.0049	0.0050
40	0. 005	0. 287	0.1154	18.89	-0.5050	13.48	152.89	50,000	760	0.0049	0.0049
45	0. 004	0. 352	0.0060	14.09	-0.0895	11.56	136.63	51,427	597	0.0044	0.0045
50	0. 018	0. 345	0.2554	16.26	-0.2230	10.58	129.88	42,732	541	0.0044	0.0045
55	0. 018	0. 358	0.3254	13.32	-0.0459	9.92	98.59	39,900	581	0.0057	0.0059
60	-0. 002	0. 324	-0.3516	18.35	-0.1420	10.58	71.93	44,429	496	0.0119	0.0120

Figure 2. 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.

Cumulative share of trading volume
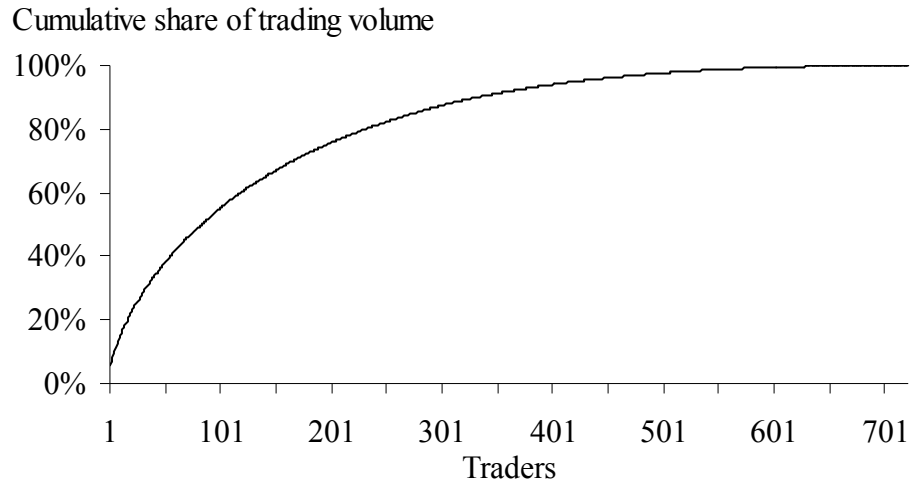
Table 2. Correlation of price impacts and trader group characteristics

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. Two additional controls in column (7) are trading volume and the number of trades over the minute just prior to a trader's trade. 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)	(7)
Trade Size	-0.015	0.026	0.135			0.042	0.021
	(-0.90)	(1.47)	(5.56)			(2.22)	(1.16)
(Trade Size)^2		-1.230	-1.748			-1.354	-1.096
		(-3.28)	(-13.24)			(-3.21)	(-3.27)
Trader Size	0.064	0.051	0.046	0.086	0.084	0.053	0.056
	(4.15)	(3.46)	(2.34)	(4.83)	(4.75)	(3.83)	(3.45)
Fin Ctr	0.133	0.129	0.137	0.154	0.168	0.128	0.125
	(5.67)	(6.33)	(4.97)	(4.07)	(4.53)	(6.35)	(6.32)
Time	-0.003	-0.002	-0.002	-0.002		-0.002	-0.002
	(10.66)	(-2.86)	(-2.25)	(-2.69)		(-2.62)	(-2.83)
Spread	0.003	0.003				0.003	0.003
	(10.66)	(13.36)				(9.58)	(12.43)
Book Vol	-0.021	-0.017				-0.019	-0.016
	(-3.89)	(-4.48)				(-4.99)	(-4.01)
Const.	0.026	0.026	0.026	0.026		0.026	0.026
	(50.41)	(49.92)	(31.21)	(26.05)		(53.72)	(49.90)
Trading volume							-0.031
							(-1.89)
Number of trades							0.009
							(1.65)
adj. R^2	0.42	0.43	0.15	0.13	0.11	0.42	0.45
AIC	-6.02	-6.07	-5.59	-5.33	-5.32	-6.41	-5.99
Het test	37.66	55.86	43.38	28.10	25.64	38.46	51.31
	(0.00)	(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	50,000

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 3. Correlation of order flow and midquote returns for different trader groups

This figure shows correlation coefficients of order flow and midquote returns on a one minute frequency for trader groups sorted by Trade Size, Trader Size, whether they come from a financial center (Fin Ctr) or not (No Fin Ctr), (trading) Time, Spread, and outstanding order book volume (Book Vol). Except for the Fin Ctr disaggregation, the x-axis shows four quartiles of traders sorted by the respective variable, where the first group consists of all traders below the 25th percentile of the cross-sectional distribution of that variable across traders, and so on. The y-axis shows correlation coefficients of a group's order flow and midquote changes.
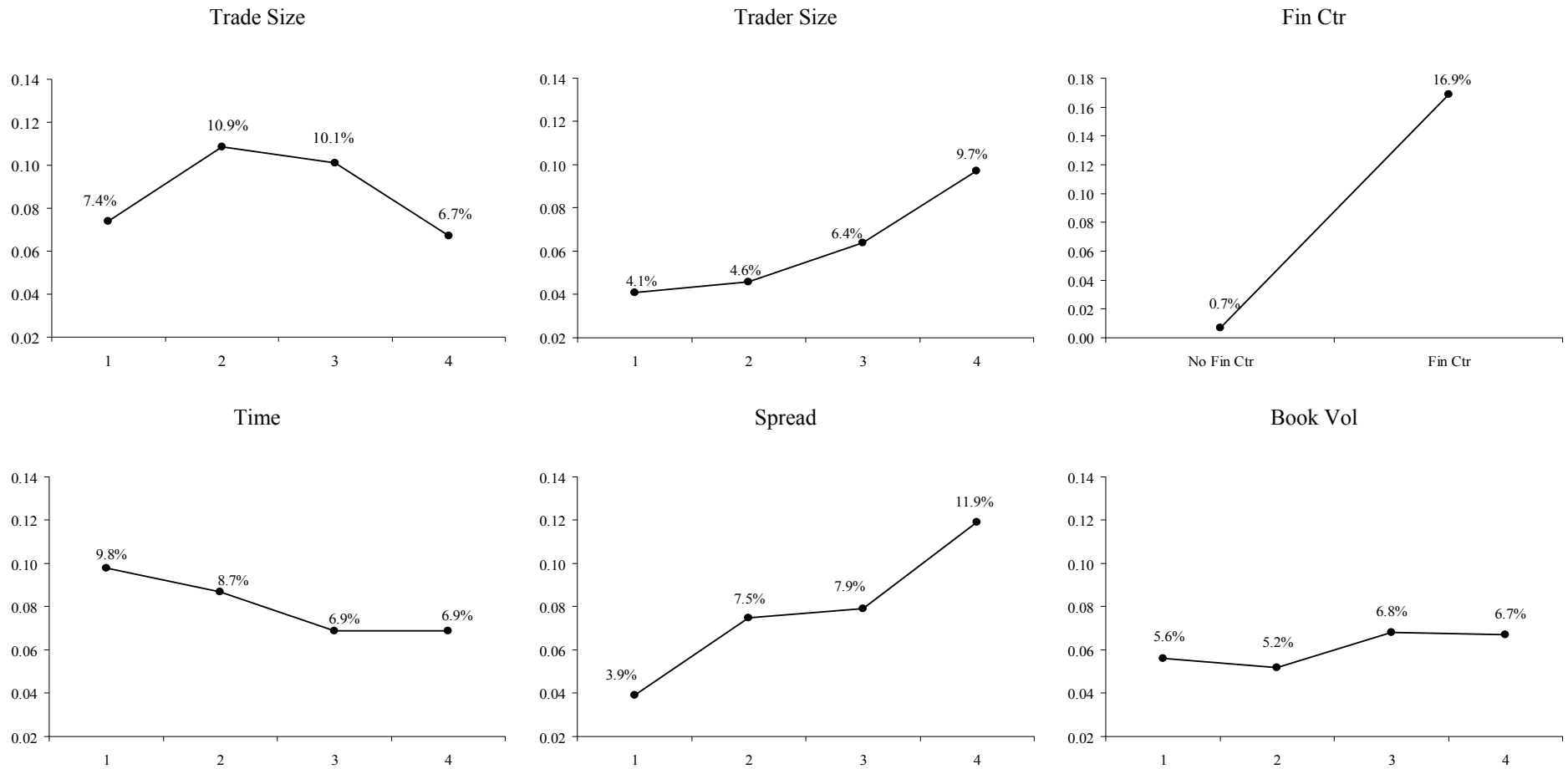
Table 5. Cross-sectional results from predictive regressions

This table shows results for regressing the t-statistic for the coefficient β_1 from the predictive regression

$$r_k^{t+1 \rightarrow t+10} = \beta_0 + \beta_{1,k} x_k^i + \varepsilon_k^{t+1 \rightarrow t+10}$$

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

	(1)	(2)	(3)	(4)
Trade Size	0.47			0.45
	(1.69)			(1.65)
Trade Size^2	-1.68	-1.78	-1.77	-1.68
	(-32.44)	(-43.60)	(-38.55)	(-38.29)
Trader Size	2.24	2.41	3.39	2.23
	(42.14)	(47.95)	(42.45)	(43.19)
Fin Ctr	2.13	1.99	1.54	2.13
	(3.99)	(3.63)	(4.84)	(3.97)
Time	-0.88	-0.95	-0.91	-0.87
	(-2.85)	(-3.14)	(-3.02)	(-2.87)
Spread	2.36	2.51	2.51	2.33
	(6.90)	(7.30)	(6.61)	(6.88)
Book Vol	-1.07	-1.12	-1.09	-1.07
	(-2.01)	(-2.15)	(-1.99)	(-2.03)
Const.	4.03	4.03	4.03	4.03
	(23.15)	(24.12)	(27.26)	(23.15)
Trading volume				-0.11
				(-1.12)
Number of trades				-0.21
				(-0.97)
adj. R^2	0.13	0.13	0.13	0.14
AIC	4.50	4.49	4.45	4.44
Obs	50,000	50,000	40,000	50,000

Table 6. Predictive regressions with time-varying coefficients

This table shows results for the predictive regression equation

$$r_k^{t+1 \to t+10} = \beta_0 + \beta_{1,k} x_k + \varepsilon_k^{t+1 \to t+10}$$

where the coefficient $\beta_{1,k}$ is a variant of the following specification:

$$\beta_{1,k} = \gamma_0 + \gamma_1 \text{Trade size}_k + \gamma_2 \text{Trade size}_k^2 + \gamma_3 \text{Trader size}_k +$$

$$\gamma_4 \text{Fin Ctr}_k + \gamma_5 \text{Time}_k + \gamma_6 \text{Spread}_k + \gamma_7 \text{Book vol} + z_k \delta$$

and z_k contains further explanatory variables used as controls for the general pace of market activity, i.e. here trading volume and the number of trades over the minute just prior to the trade. T-statistics in brackets are based on HAC Newey-West standard errors. Specifications (1) to (5) are based on the whole sample, whereas specifications (6) and (7) are estimated on the first 15 minutes and the next 35 minutes, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
x (order flow)	3.47 (4.24)	3.43 (4.17)	3.43 (4.34)	3.40 (4.08)	3.39 (4.07)	3.69 (4.21)	3.25 (4.70)
Trade Size			0.06 (1.01)				
Trade Size^2			-0.29 (-2.91)	-0.29 (-2.79)	-0.27 (-2.78)	-0.22 (-2.01)	-0.29 (-2.34)
Trader Size			0.27 (2.10)	0.29 (2.43)	0.29 (2.44)	0.36 (2.77)	0.25 (2.12)
Fin Ctr			0.45 (3.20)	0.45 (3.20)	0.44 (3.19)	0.35 (2.57)	0.43 (3.01)
Time			-0.19 (-2.21)	-0.24 (-2.79)	-0.25 (-2.87)	-0.14 (-1.99)	-0.24 (-2.68)
Spread			0.48 (4.07)	0.59 (4.89)	0.51 (4.12)	0.61 (5.25)	0.50 (3.59)
Book Vol			-0.25 (-2.32)	-0.37 (-2.83)	-0.30 (-2.61)	-0.38 (-2.91)	-0.31 (-2.19)
Trading volume		-0.09 (-3.55)	-0.06 (-2.35)		-0.06 (-2.34)	-0.05 (-1.95)	-0.09 (-2.22)
Number of trades		0.02 (1.30)	0.00 (0.47)		0.01 (0.97)	0.00 (0.57)	0.02 (1.15)
Const.	3.25 (3.92)	2.28 (1.10)	2.30 (1.09)	2.29 (1.09)	2.30 (1.09)	4.01 (4.34)	-1.21 (0.81)
Adj. R^2	0.03	0.05	0.11	0.09	0.10	0.08	0.13
AIC	10.03	10.00	9.68	9.73	9.72	10.17	9.45
Obs	12,972	11,889	11,889	12,972	11,889	6,308	5,581

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
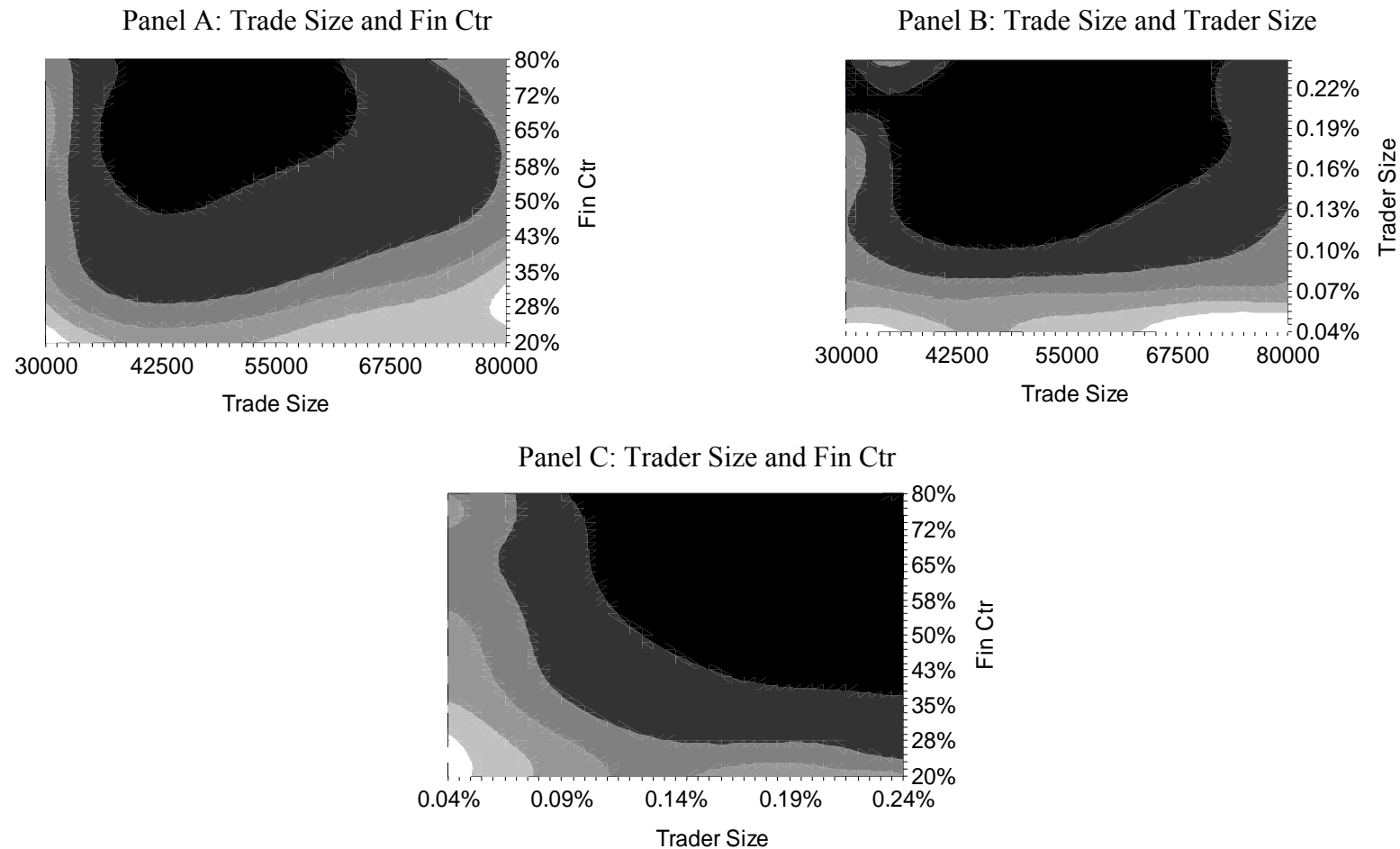
Appendix 1. Descriptive statistics for simulated trader groups

This table shows descripticve statistics for price impacts (PI) and the six trader group charac-
teristics: (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