Lack of Anonymity and the Inference from Order Flow

Juhani Linnainmaa and Gideon Saar

This version: September 2010

Juhani Linnainmaa is from the University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637 (Tel: 773-834-3176, jlinnain@chicagobooth.edu). Gideon Saar is from the Johnson Graduate School of Management, Cornell University, 455 Sage Hall, Ithaca, NY 14853 (Tel: 607-255-7484, gs25@cornell.edu). Henri Bergström from the Finnish Central Securities Depository and Pekka Peiponen, Timo Kaski, and Jani Ahola from the Helsinki Exchange provided data used in this study. Matti Keloharju provided invaluable help with the FCSD registry. We are grateful for comments from Ekkehart Boehmer, Ronald Goettler, Søren Hvidkjær, Robert Jarrow, Tobias Moskowitz, Ioanid Rosu, and seminar participants at Cornell University. Linnainmaa gratefully acknowledges financial support from the University of Chicago Booth School of Business.

Lack of Anonymity and the Inference from Order Flow

Abstract

This paper investigates the information content of signals about the identity of investors and whether they affect price formation. We use a dataset from Finland that combines information about the identity of investors with complete order flow records. While we document that investors use multiple brokers, our study demonstrates that broker identity can nonetheless be used as a powerful signal about the identity of investors who initiate trades. This finding testifies to the existence of frictions in the economic environment that prevent investors from completely eliminating the information content of broker ID using mixed strategies. We show that the broker ID signal is important enough to affect prices: The permanent price impact of orders coming from different brokers fits the information profile of the investors associated with these brokers. Our results suggest that the market correctly processes the signal embedded in broker identity, and liquidity improvements documented in the literature when exchanges adopt a more anonymous market structure arise because prices adjust less efficiently to order flow information when the degree of anonymity increases.

1. Introduction

The concept of anonymity in financial markets pertains to the information market observers have about the identity of investors who submit orders. On the most basic level, anonymity relates to how informed investors exploit their informational advantage, and therefore the extent of anonymity affects adverse selection and the informational efficiency of prices. As markets around the world respond to competitive and regulatory pressures, features of market design such as the rules governing disclosure of trader information (and transparency in general) receive considerable attention. Understanding how the building blocks of market design affect the strategies of traders, the evolution of prices, and the survival of competing execution venues is of great interest to a wide range of market participants and researchers.

The extent of anonymity associated with distinct market structures can be measured along a spectrum: On one end there are brokered upstairs markets in which the identity of investors is often shared with trading counterparties, in the middle are electronic limit order books that display broker identity as well as floor markets, and on the other end are electronic limit order books that disclose no information about incoming orders beyond size and price.[1] Several recent event studies find that liquidity can change when the degree of anonymity in the market is altered. More specifically, these studies look at what happens when exchanges begin or stop displaying the broker identity (henceforth, ID) associated with orders.[2]

Why is broker ID important? The usual explanation is that if different brokers represent separate clienteles with distinct attributes, the order flow handled by a broker will represent a mix of informed and uninformed order flow that could differ from the unconditional mix of

[1] Traditional floor markets are similar to electronic limit order books with displayed broker ID in that the identity of the broker is known to floor traders. The floor also fosters personal relationships among brokers and other intermediaries (e.g., Designated Market Makers on the NYSE) that often facilitate additional sharing of information about the identity of investors (see, for example, Benveniste, Marcus, and Wilhelm (1992), Battalio, Ellul, and Jennings (2007), and Garfinkel and Nimalendran (2003)).

[2] See, for example, Comerton-Forde, Frino, and Mollica (2005), Foucault, Moinas, and Theissen (2007), Aspris, Frino, Gerace, and Lepone (2008), Maher, Swan, and Westerholm (2008), and Majois (2009). The first three papers find that increased anonymity improves liquidity, though the last two papers claim that this result could be sensitive to the econometric specifications employed.

orders in the market. Therefore, the identity of the broker could potentially be used as a (noisy) signal about the identity of the investor. It is unclear, however, why liquidity in the market should be affected by the availability of the broker ID signal. Informed investors could use mixed strategies (i.e., trade through multiple brokers) in a manner that makes this signal completely uninformative. In the absence of frictions in the economic environment, the profit maximizing strategy of the informed investors would be to hide using multi-broker trading, rendering the market structure with an intermediate level of anonymity irrelevant.

Our goal in this paper is to investigate the information content of signals about the identity of investors and whether they affect price formation. By that, we hope to shed light on whether frictions in the economic environment create and sustain such informative signals, and have a better understanding of why changes in market design that are aimed at altering the level of anonymity affect liquidity. We utilize for this purpose a dataset from Finland that combines information about the identity of investors with complete order flow records.

We start by grouping investors into three types: households, domestic institutions, and foreigners. Our maintained assumption throughout the analysis is that households are less likely than domestic institutions to possess valuable private information. This assumption gives us the ability to identify informed and uninformed investors ex-ante and to use that knowledge to test the market's inference from broker ID.[3] We observe that 11.5% of the domestic institutions and 9.5% of the households use more than one broker on at least 25% of the days in which they trade, but these multi-broker users generate 81.0% (20.7%) of the trading by domestic institutions (households). The fact that most trading by domestic institutions utilizes brokers in this manner suggests that hiding is important and institutions are aware of the information content of their order flow (as well as the market's attempt to infer information from broker ID).

[3] Foreigners constitute a more ambiguous case in that it is difficult to say ex-ante whether they are more or less informed than households or domestic institutions: Most foreigners that are active in Finland are probably sophisticated foreign institutions, but their distance from the market could give locals an advantage. We therefore report the results for foreigners as stylized facts rather than use them to test the market's inference.

We therefore ask whether broker ID can nonetheless be used to extract meaningful signals about the types of investors who initiate trades even when investors utilize multiple brokers. To facilitate the empirical analysis, we use a very simple rule to categorize brokers: A broker belongs to a Broker Group that is associated with a certain investor type (households, foreigners, or domestic institutions) if more than 50 percent of the trades that the broker executes involve this investor type. We show that considering broker group identifiers alongside other explanatory variables markedly increases our ability to say whether an order that initiates a trade comes from a household, a domestic institution, or a foreigner. Specifically, we use probit models to classify the investors who initiate trades given the attributes of the orders they send and the market conditions when the orders arrive.[4] The pseudo-R^2 of the pooled probit model increases when we add the broker group dummies in the households regression from 25.06% to 43.95%, and it similarly increases in the foreigners regression from 5.66% to 30.45% and in the domestic institutions regression from 2.58% to 15.77%. We therefore conclude that the extent of multi-broker usage in the market is insufficient and broker ID is informative about investor types.

We then proceed to investigate whether this information is important enough to affect prices. In other words, we look for evidence in prices that is consistent with the use of broker ID by market participants to infer information about the identity of investors. We find that the permanent price impact of marketable orders coming through brokers associated with households is smaller than that of orders coming through brokers associated with domestic institutions (and foreigners). Across subsamples featuring varying degrees of information asymmetry, the permanent price impact of marketable orders coming from brokers associated with domestic institutions increases with the extent of information asymmetry, while the permanent price

[4] Investors initiate trades by submitting marketable (limit) orders. These orders are priced in a manner that triggers immediate execution upon arrival to the market and hence do not enter the book. The Helsinki Stock Exchange, like most electronic limit order books, requires every order to have a limit price. Therefore, these orders are called marketable (limit) orders to differentiate them from market orders in traditional dealer and floor markets in which the order does not specify a price.

impact of orders coming from brokers associated with households does not. These two findings suggest that market participants use broker ID to make inferences about the prevalence of price-relevant private information in the order flow. Orders that are attributed to less informed investors are afforded with a lower permanent price impact that does not change much with the degree of adverse selection in the environment, while the opposite can be observed for orders that are inferred to come from more informed investors. We use various specifications to show that these results hold when we account for order attributes and market conditions that can also be used to make an inference about the identity of investors.

We further find that the inference made by investors is rather sophisticated in that other attributes of the order are used in conjunction with broker ID to refine the signal about investor types. For example, without conditioning on the size of arriving marketable orders, 72.3% of the orders that come from the broker group that we associate with households are indeed household orders. When we examine the order flow of this broker group by order size, 81.7% of the smaller orders but only 40% of the larger orders come from households, which means that the larger orders they receive are more likely to come from domestic institutions than from households.

Our regression results show that these considerations are reflected in prices: The interaction of trade size with the dummy variable for the broker group that is associated with households is positive, consistent with the inference that larger trades are more likely to come from informed investors. On the other hand, the interactions of trade size with the dummy variables for broker groups that we associate with domestic institutions and foreigners are negative, consistent with Barclay and Warner (1993)'s stealth-trading hypothesis whereby informed investors split large orders into medium-sized chunks. We find interesting interactions of broker identity with other characteristics of the economic environment such as the prevailing spread in the book when the order arrives and time from the last trade. These findings mean that the market pays attention to details: If a broker that is associated with households executes a trade that deviates considerably from what its "usual" trade looks like, the price impact can be greater than if the trade originated from a broker associated with domestic institutions.

One contribution we make is that our findings explain why empirical investigations of events in which the degree of anonymity changes (either introducing or eliminating the display of broker ID) show a significant impact on market liquidity. The typical result (e.g., Comerton-Forde, Frino, and Mollica (2005), Foucault, Moinas, and Theissen (2007), and Aspris, Frino, Gerace, and Lepone (2008)) is that liquidity improves when broker ID is eliminated. We find that the market uses broker ID information to update prices. Without this signal, informed investors would be able to hide more effectively. A typical information asymmetry model would suggest that the price impact of a trade would decrease in this case (i.e., that liquidity would improve) but at the expense of informational efficiency: Prices simply do not adjust sufficiently to the order flow, enabling informed investors to realize greater profits from their informational advantage. The greater profits of the informed investors would likely come at the expense of the uninformed investors, and hence it is not obvious that the finding of improved liquidity in the market is beneficial to society at large. Furthermore, by making stock prices less efficient, anonymity creates a negative externality in that corporate decisions are based on less-efficient prices and derivatives on these stocks are more difficult to value.[5] Hence, our findings lead to a somewhat less positive interpretation of the result in the event study literature that liquidity improves when broker ID is removed from the market.

Models that study anonymity often feature a signal about the identity of some traders (e.g., knowledge about the demand coming from uninformed investors) and analyze how the existence of this signal affects the strategies of market participants and the evolution of prices.[6] An important assumption that underlies these theoretical models is that investors utilize signals about the identity of those submitting orders and that prices adjust accordingly. Our finding that inference about broker ID affects the permanent price impact of orders is consistent with such

[5] The negative impact of broker ID removal on informational efficiency may not occur in a framework where endogenous information acquisition is taken into consideration. However, this would likely depend on specific parameter values rather than be a more general result (see Rindi (2008)).

[6] See, for example, Röell (1990), Admati and Pfleiderer (1991), Benveniste et al. (1992), Foster and George (1992), Madhavan (1996), Battalio and Holden (2001), Frino, Johnstone, and Zhang (2005), Foucault, Moinas, and Theissen (2007), Boulatov and George (2008), and Rindi (2008).

optimal behavior by market participants. Still, we do not observe that informed investors utilize multiple brokers to the extent that makes broker ID uninformative, as would be prescribed by adverse selection models without additional frictions. Therefore, another contribution of our study is that we document evidence that is consistent with the existence of powerful frictions in the economic environment that prevent investors from using multiple brokers "optimally" to hide their identity. These frictions could be driven by bundled brokerage services, commission discounts, reputation, or bounded rationality, and they sustain the value of the signal that can be extracted from broker ID. Our work complements the analysis of Goldstein, Irvine, Kandel, and Wiener (2009) who look for evidence on the existence of such frictions by studying institutional trading patterns and commission rates.

The rest of this paper proceeds as follows. Section 2 presents our data and sample. Section 3 discusses the mapping we create from broker identities to investor types, shows the extent of multi-broker use by investors, and demonstrates that broker ID can be used as an informative signal about the investor type behind orders. Section 4 looks at how the inference from broker ID affects prices, beginning with simple comparisons and continuing with more structured models that control for order attributes, characteristics of the limit order book, and recent market activity. Section 5 presents our conclusions.

2. Data and Sample

2.1. Data

Our first data source is the Finnish Central Securities Depository (FCSD) registry, which contains the complete trading records of all Finnish investors in publicly-traded domestic stocks. Each trade record includes the date, price, volume, and a buy/sell indicator. The data also provide certain details about the investor behind each trade, including a categorization of investors into various types (households, non-finance corporations, finance and insurance corporations, government, non-profit institutions, and foreigners).[7] For the purpose of this study, we group all investors into three "investor types": Households, domestic institutions (i.e., all domestic investors except households), and foreigners.[8]

Our second data source is Helsinki Stock Exchange (HEX) supervisory files, which contain detailed information on every order that is entered into the system, including date and time, the broker submitting the order, price, number of shares, and a set of codes that enables us to track the order throughout its life (from submission to execution, modification, or cancellation). The data can be used to completely reconstruct both the second-by-second limit order book and the record of executions for all stocks traded on the exchange.

Trading on the HEX starts every day with an opening call auction and proceeds with a continuous trading session utilizing an electronic limit order book market structure. The HEX electronic limit order book follows the standard price-time priority rules according to which the best-priced orders in the book are executed first (e.g., lower-priced sell orders) and, within a price level, orders that arrive at the exchange first are executed before orders that arrive later. Investors who observe the limit order book can see the Broker IDs of standing limit orders. A separate real time "trade feed" ensures that all investors (including individual investors who use online brokers) observe the broker ID of executed orders.

2.2. Matching

Our study requires that we determine which investor type is behind each trade, and therefore we need to match the HEX data with the FCSD registry data. The matching algorithm starts by identifying all trades in the two datasets for which there is an exact match using stock, date, price, and volume information. Since the HEX data structure enables us to track a single order through multiple partial executions, it is enough to have an exact match that identifies one of these executions to be able to complete the necessary information for the rest. The fact that we

---

[7] See Grinblatt and Keloharju (2000) for a detailed exposition of the FCSD registry.

[8] Most foreigners choose to trade through nominee accounts handled by domestic financial institutions. A special indicator enables us to identify their trading, and we use the term "foreigner" throughout this study to refer to those foreigners who trade through nominee accounts.

group investors into three types (households, domestic institutions, and foreigners) helps the matching procedure because we do not need to match each trade to a specific investor. Rather, if there is an ambiguous case but our matching procedure results in two candidate investors from the same investor type, we can safely classify the investor behind the order as belonging to that investor type.[9]

Our matching procedure successfully identifies the investors behind most trades: We unambiguously match 90.5% of the trades in the average stock each day. Figure 1 provides additional information on the stability of the matching rate. To construct this figure we compute the daily matching rate (the number of trades for which we have identified the investor type behind the trade divided by the total number of trades) for each stock and take the equal-weighted average to find the average matching rate for the day. We then plot the daily matching rate over our sample period. We observe that the matching rate is very stable over time, which increases our confidence in the identification procedure.

2.3. Sample

Our sample period starts on July 10, 2000 and ends on October 23, 2001. Our universe of securities includes all stocks that were listed on the HEX during the sample period (i.e., we exclude warrants, employee stock options, and temporary share classes that are traded on the exchange). We compute the number of trades per listed firm, and retain a listed firm in the sample if it has at least five trades per day on average during our sample period.[10] This trading activity screen results in 87 listed firms that constitute the sample.[11] Table 1 reports summary statistics for the sample firms. Our analysis uses all 41 brokerage houses that execute trades in these stocks during the sample period.

---

[9] A detailed explanation of the matching algorithm is available from the authors upon request.

[10] If a listed firm has multiple share classes, we aggregate across share classes for the purpose of computing the daily average number of trades.

[11] Certain listed firms have multiple share classes, and therefore our sample of 87 firms has 109 separate ticker symbols that are traded in the market. Different share classes of the same firm tend to differ from each other in terms of liquidity as well as the level of institutional ownership, and therefore we keep each share class separate in our analysis.

3. Investor Types and Broker Identity

3.1. Investor Types and Broker Groups

Our empirical analysis studies three, well-defined investor types: (i) all domestic households, (ii) foreign investors who trade via nominee accounts, and (iii) domestic institutional investors.[12] The latter investor type is an umbrella category that includes all domestic investors other than households (i.e., finance and insurance institutions, non-financial corporations, nonprofits, and government). While it is possible that the trading of each group has its own characteristics, the consideration that guided our decision to lump them together is that we could not identify specific brokerage houses that predominantly serve only one of these subsets of domestic institutions (the amount of trading of each subset is not very large and brokerage houses that cater to domestic institutions serve all of them). When lumped together, however, it is indeed possible to find brokerage houses that do most of their business with domestic institutions.

We acknowledge that there is probably much heterogeneity in terms of skill or sophistication within this group. Nevertheless, the extent of price-relevant information, skill, or sophistication of the average domestic institution in this group should be greater than that of the average domestic household. Therefore, the maintained assumption that underlies our tests of inference from broker ID is that institutions are more "informed" than households.[13] As for foreigners, we are unsure ex-ante whether they are more or less sophisticated than households or domestic institutions.[14] Therefore, the empirical evidence we gather serves to document the market's beliefs about the foreigners' skill level rather than to test any hypothesis about anonymity.

---

[12] Since we do not have information about the foreign entities that trade via nominee accounts, this category could potentially include both individual and institutional foreign investors. Our "foreigners" investor type is therefore a mixture (with unknown weights) of these two investor groups.

[13] Grinblatt and Keloharju (2000) use Finnish data from 1995 and 1996 to look at a measure of 6-month performance of different investor types. They conclude that the performance of institutions is better than that of individuals, which is consistent with our maintained assumption that institutions are more informed than individuals.

[14] There are various arguments in the literature as to whether foreigners are more sophisticated and hence have an advantage in discovering and trading on information relative to domestic investors, or whether they are at a disadvantage due to their distance from the firms' headquarters and primary markets. See, for example, Shukla and van Inwegen (1995), Grinblatt and Keloharju (2000), and Choe, Kho, and Stulz (2005).

How do investors know the mapping from broker identities to investor types? It could be "common knowledge" among market participants that certain investor types in specific stocks trade through particular brokers. It is also possible that sophisticated investors mine datasets similar to the one we use in this paper to learn more about this mapping. While it is difficult to ascertain how fine a mapping market participants use, we implement a very simple rule to facilitate the empirical investigation in this paper: A broker belongs to a Broker Group that is associated with a certain investor type (households, foreigners, or domestic institutions) if more than 50 percent of the trades that the broker executes involve this investor type. We apply this rule on a stock-by-stock basis because some brokers could see the majority of their orders coming from one investor type (e.g., domestic institutions) in large and active stocks, while having most of their trading in thinly traded stocks come from another investor type (e.g., households). Since price inference is predominantly done on a stock-by-stock basis, market observers can easily take into consideration the characteristics of the stock in addition to the broker identity.

Our rule that assigns brokerage houses to three investor types slightly reduces the average number of brokers per stock, from 26.98 to 24.84.[15] The assigned brokers execute nine out of ten trades in the market: The cross-sectional average (median) ratio of assigned trades to all trades in the market is 89.1% (92.4%). Hence, our rule leaves the majority of trades in the sample, and has the advantage that it is simple and unambiguous.[16] By definition, however, many trades coming through the assigned brokers in each broker group are incorrectly classified in that they are initiated by one of the other two investor types. This introduces noise into the analysis and

---

[15] We lose a few brokers that serve all three investor types more or less equally and therefore cannot be assigned to a broker group.

[16] We assign brokers to broker groups based on trading over the entire sample period. However, we also examined the persistence of broker assignments to verify that brokers' customer bases do not change often, which would frustrate attempts by market participants to extract a useful signal from broker ID. Specifically, we assign brokers to the broker groups based on executed trades each month (if they execute at least 100 trades per month), and examine whether the same assignment is observed in the following month. The results are very encouraging: For example, if a broker is assigned to the broker group associated with households in month *t*, the likelihood that it retains this assignment in month *t+1* (rather than being assigned to the other two investor types) is over 96%.

should prevent us from observing differences among the broker groups if the rule we use is not effective enough.

Table 2 provides information about the extent of this noise. We report the fraction of trades (and separately of volume) in each broker group that comes from the actual investor types.[17] We observe that for Broker Group 1 (henceforth, BG1), which we associate with households, 72.3% of the trades originate from households, while 20.8% come from domestic institutions, and 6.9% come from foreigners.[18] For Broker Group 2 (henceforth, BG2), which we associate with foreigners, 79.1% of the trades come from foreigners, 17.2% come from domestic institutions, and 3.7% come from households. Finally, 71.6% of the trades of Broker Group 3 (henceforth, BG3), which we associate with domestic institutions, originate from domestic institutions, 11.5% come from households, and 16.9% come from foreigners.

These numbers demonstrate that our broker assignment rule results in significant differences between the fraction of trades associated with the dominant investor type for each broker group and the fractions of trades coming from the other investor types. This is a key element in using broker ID for making an inference about investors. It is important to stress that we later test the pricing implications of inference from broker identifies, not how prices respond to the trading of particular investor types. This is why we label these groups BG1, BG2, and BG3 rather than utilizing the name of an investor type to represent them. Each broker group executes orders of all investor types in various quantities. Investors can only observe the broker ID (and hence know the broker group), not the identity of the investor behind the broker ID.

3.2. Use of Multiple Brokers

A simple sequential trade model with information asymmetry could be used to show that without any frictions associated with broker selection, informed investors would want to use mixed

[17] We provide two versions of this measure: (i) first computing the fraction for each firm and then taking an equal-weighted average to create a cross-sectional statistic (EW), and (ii) first aggregating the trades of the brokers across all the stocks and then computing the fraction for each broker group (UW). The first definition takes the perspective of a “typical” stock, while the second one takes the perspective of a “typical” trade.

[18] The fraction of volume coming from households through the brokers we assign to BG1 is somewhat smaller than the fraction of trades because households execute smaller trades on average.

strategies across the different brokers. More specifically, they would trade such that the information content (and hence the price impact) of an otherwise identical order sent through two different brokers is the same. If brokers differ from each other with respect to the amount of uninformed order flow they handle, the optimal amount of informed order flow would also be different. If the price impact of an otherwise identical order is not the same across brokers, informed investors would have an incentive to switch brokers and hence equilibrium (absent additional frictions) would require that we observe multiple broker use.

We therefore look at whether specific investors (e.g., a particular household or financial institution) use multiple brokers. Since we cannot observe a unique identifier for each foreign investor (because they trade via nominee accounts), this analysis is restricted to domestic households and institutions. In the first line of Table 3 we define a "multi-broker user" to be an investor who uses more than one broker on at least 25% of the days in which it trades. We observe that 9.1% of the households satisfy this definition, as well as 11.5% of domestic institutions. These individuals and institutions, however, generate a disproportionate amount of the volume in the market. In particular, the institutional multi-broker users are responsible for 81.0% of all trading by domestic institutions! The second line in the table defines a multi-broker user as an investor who uses multiple brokers in the same stock during the same week.[19] We observe that the fraction of investors who are categorized as users of multiple brokers remains remarkably similar: 8.1% for households and 12.0% for domestic institutions, with 17.3% and 68.1% of the trades of these two investor types, respectively.

This evidence, while striking, does not necessarily imply that informed investors are trying to optimally hide their trading. The reason is that if an investor trades through multiple brokers that belong to the same broker group, the realized benefits from attempting to camouflage the order flow are lower compared to when the investor trades via brokers that

[19] Chan and Lakonishok (1995) note that many institutional investors split a larger order into a sequence of smaller orders (that they term a "package") and trade them over multiple days: More than half of the dollar value traded in their sample consists of packages that take four or more days to completely execute.

belong to different broker groups. The last two lines of Table 3 define multi-broker users as those that trade the same stock via brokers that belong to at least two different broker groups either in the same week or on the same day. The percentage of investors who do that falls considerably: 1.6% of individuals and 3.8% of institutions trade the same stock using multiple broker groups on the same week. Nonetheless, these domestic institutions are responsible to more than half of total trading by institutions.

We emphasize a couple of observations from Table 3. First, the percentage of trading that comes from multi-broker users is much higher among domestic institutions than among individuals. While certain features of the brokerage business could drive this pattern, it is also consistent with a greater need for "hiding" by domestic institutions and hence a greater likelihood that they are informed investors, consistent with our maintained assumption. Second, what matters for pricing in traditional sequential trade models is the probability that a trade comes from an informed investor. Therefore, the fact that most institutional trades come from multi-broker users suggests that hiding is important; active institutions are aware of the information content of their order flow and would like to frustrate the market's attempt at inferring information from their trades.

3.3. Broker ID as a Signal

We continue our investigation by asking whether broker ID contains meaningful information about the types of investors who initiate trades even though investors utilize multiple brokers. In general, there are various attributes of trades and the trading environment that can help market participants infer who is behind a trade. The real question is whether broker ID can assist in making such an inference over and above these attributes. To evaluate this question we run a probit regression for each investor type. In the regression for households, the dependent variable is set to one if a household submits a marketable order (i.e., an order that results in a trade) and zero if either a foreigner or a domestic institution submits the order.

The explanatory variables consist of broker group dummies (for the three broker groups) and a set of variables that could also potentially be used to make an inference about the investor behind the order. The set of variables includes recent activity (volume, signed return, and volatility in the previous five minutes; log duration from the previous trade and log duration squared), the prevailing state of the limit order book (same-side and other-side depth at the best price; log prevailing spread and squared log spread) and trade size variables (log and squared log of trade size; a dummy variable for trades that are equal to the displayed depth at the best price in the book, and a dummy variable for marketable orders that are larger than the displayed depth).[20] The probit regressions are estimated using observations pooled from all stocks and include stock-specific fixed effects (which are omitted from the table for ease of presentation).[21]

Table 4 presents two regressions for each investor type: The first includes only the broker group dummies, and the second adds the other explanatory variables. We observe that adding the explanatory variables for order attributes and market conditions does not change our conclusions. In the households regression, the coefficient on the dummy variable for BG1 is positive and significant, which means that observing a marketable order from a BG1 broker increases the likelihood that this order originated from a household. On the other hand, the negative coefficients on BG2 and BG3 indicate that marketable orders from these brokers are associated with a lower likelihood of coming from a household. In the regressions for foreigners in the third and fourth columns, the dependent variable is set to one if a foreigner submits the marketable order and zero if either a household or a domestic institution submits the order. We observe an analogous pattern: The coefficient on BG2 (the broker group we associate with foreigners) is positive and highly significant while the coefficients on BG1 and BG3 are negative. A similar pattern is observed in the fifth and sixth columns where we analyze domestic institutions.

---

[20] Volume is defined as turnover (the number of shares traded divided by the number of shares outstanding). Our volatility variable is the logarithm of the ratio of highest to lowest trade prices in the previous five minutes. Duration is defined as log[1+ time from the previous trade in seconds], and trade size is defined as log[number of shares traded * price].

[21] One way with which to view this test is that we assume the market knows the true model linking trade attributes to the underlying investor, and we measure how powerful a statistical model aimed at distinguishing the three investor types can be. This analysis is analogous to spanning tests from the asset pricing literature that look at whether one set of factors subsumes another set of factors. Here, the question is whether the market should use broker identities in addition to (or instead of) other order and market environment attributes to make an inference about the type of investor that is behind a trade.

Overall, the probit regressions show that broker group identities, as captured by our classification scheme, help reveal the identity of the underlying investor. How much information is there in broker ID? The pseudo-R^2 of the pooled households regression is 25.06% when all explanatory variables except broker group dummies are included in the regression, and it substantially increases to 43.95% when the broker group dummies are included. If we add a separate dummy variable for each of the 41 brokers in our sample rather than the dummies for the three broker groups, the pseudo-R^2 in the households regression increases to 46.66%. This means that our classification into broker groups is almost as effective as knowing the specific identity of the broker through which the order is submitted. Similarly, we observe a large increase in pseudo-R^2 when we add broker group dummies to the foreigners regression (from 5.66% to 30.45%) and to the domestic institutions regression (from 2.58% to 15.77%).[22]

Beyond demonstrating the strength of the broker identity signal, the results in Table 4 already provide some information about how the trading strategies of households, foreigners, and domestic institutions differ from each other. For example, the coefficient on the prevailing spread is negative in the households regression while it is positive in both the foreigners and the domestic institutions regressions. This indicates that the latter two investor types tend to trade when spreads are wide, perhaps because they have a greater incentive to trade when there is more information asymmetry in the environment. The coefficient on duration indicates that domestic institutions are more likely to send marketable orders when there is an increase in the pace of trading, which again could signal an information-intense period, while households do the opposite.

The coefficient on trade size is positive in the regressions for all three investor types, contrary to our expectation that it would be negative for households because they tend to submit smaller orders than foreigners or domestic institutions. To further investigate this result, we

[22] We also ran the probit regressions for each stock separately rather than using pooled regressions with stock fixed effects. The average pseudo-R^2 of the households regressions increased from 25.80% to 44.74% when the broker group dummies were added. The average increases for the foreigners and domestic institutions regressions were from 6.37% to 30.29% and from 2.52% to 16.02%, respectively.

divide the sample of marketable orders into quartiles according to their size and run separate probit regressions that include stock fixed effects and the same set of explanatory variables as in Table 4 for each trade-size quartile.[23] Figure 2 presents the BG1 coefficients from the households regressions in each trade-size quartile, as well as the BG2 coefficients from the foreigners regressions and the BG3 coefficients from the domestic institutions regressions.[24]

The coefficient on BG1 in the households regression is positive in the smallest trade-size quartile, but decreases and becomes negative for the third and fourth quartiles. This means that observing a large marketable order sent through a broker that predominantly serves households actually decreases the likelihood that it comes from a household investor. This finding motivates looking at interactions of broker groups and other explanatory variables (especially trade size) in our price impact analysis. Such effects are not observed in the regressions for the other two investor types: The coefficient on BG2 in the foreigners regression is positive and is of the same order of magnitude across the trade-size quartiles, and the coefficient on BG3 in the domestic institutions regression is positive in all quartiles as well. It is interesting to note, though, that the coefficient on BG3 increases in magnitude considerably when we move from the smallest to the largest trades, which means that our ability to correctly attribute to domestic institutions a trade that comes through a BG3 broker is much better when the trade is larger.

In summary, the findings documented in this section show that while much of the trading (especially of institutional investors) come from multi-broker users, broker ID can nonetheless be used by market participants to significantly increase their ability to infer who (in terms of our investor types) is behind initiated trades.

---

[23] The breakpoints for the trade-size quartiles are stock specific, and therefore each stock contributes observations to all trade-size quartiles.

[24] All coefficients presented in Figure 1 are statistically significant at the 1% level.

4. Broker ID Information and Prices

Given the result of the previous section that broker ID contains information, a natural question is whether market participants use broker ID to make an inference about investor types. In particular, we look for evidence in prices to tell us whether broker ID information is meaningful enough to affect price formation in the market.

4.1. Permanent Price Impact Tables

The primary variable with which we investigate the collective inference about investor types the market makes from broker ID is the permanent price impact of trades. A trade occurs when a standing limit order is executed by an incoming marketable order. Hence, the new information brought to the market at the time of the consummation of the trade is the arrival of the marketable order that initiated the trade. Suppose that market participants know the investor type with certainty after observing each trade. If they are certain that the investor who just executed a marketable order is uninformed, the posterior beliefs about the true value of the asset remain unchanged. On the other hand, if market participants are certain that it was an informed investor who just initiated a trade, their beliefs about the asset's true value should adjust to reflect this knowledge. The adjustment in this case depends on the market's belief about the value of the information that the informed investor possesses, but the true value must be at least as high (low) as the ask (bid) price if the informed investor bought (sold) shares.[25] As broker ID provides a (noisy) signal about the probability that an investor is either informed or uninformed, we look for evidence in the permanent price impact of trades for the utilization of this signal.

The permanent price impact of a trade (or a marketable order) attempts to measure price adjustment from an instant before the arrival of the marketable order to a time where we assume prices have finished their adjustment to the information content of the order.[26] Researchers in the

---

[25] Of course, if informed investors have private hedging needs or are subject to liquidity shocks, their trading strategy would reflect the combined influences of these considerations and hence inferences about the asset's true value from their trading would be more complex.

[26] Market microstructure often differentiates between the temporary price impact of a trade (sometimes called the realized spread) that is measured from the trade price to a post-trade benchmark price, and the permanent price impact that is measured from a pre-trade benchmark price to a post-trade benchmark price. The temporary price

market microstructure literature often use the prevailing midquote (i.e., the midpoint between the bid and ask prices) as the pre-trade benchmark price (see, for example, Hasbrouck (2009)). Because there is no one price for a security in a limit order book, the midquote prevailing at the time an order arrives seems like a natural choice because it is a single price that can be used as a benchmark for both buys and sells. Similarly, after the market adjusts to the information content of the order through a combination of cancellation and modification of existing limit orders, execution of additional marketable orders, and submission of new limit orders to the book, the midquote should again be a useful benchmark for the value of the security.

How long it takes for the adjustment is difficult to judge. Market microstructure studies often use the midquote five minutes after a transaction as a post-trade benchmark price, a convention we adopt here as well though we also carry out robustness tests using other definitions.[27] Table 5 presents the permanent price impact of marketable orders in several subsamples. We believe it is instructive to look first at some preliminary numbers without imposing additional structure, and postpone the regression analysis that controls for trade attributes and the prevailing market conditions to the next section. The definition of permanent price impact that we use is the signed log change in the midpoint of the quote from an instant before the transaction to five minutes after the transaction. In order to account for cross-sectional heterogeneity in permanent price impacts, we standardize the measure by subtracting from each observation the average price impact in the same stock for marketable orders in the same direction (i.e., separate averages are computed for buys and for sells). To facilitate interpretation, all price impacts are multiplied by a factor of 100. Therefore, a value of "1" indicates that the standardized log-change in the quote midpoint is 1%.

---

impact is usually ascribed to inventory costs (or in general to risk-averse liquidity provision) or order processing costs, while the permanent price impact is ascribed to fundamental private information along the line of models such as Glosten and Milgrom (1985) or Kyle (1985). The total price impact of a trade (a measure of trading costs) is the sum of the temporary and the permanent price impacts.

[27] Using the midquote five minutes after the trade as a post-trade benchmark price became even more commonplace when the SEC adopted it as a standard for the computation of execution costs and required U.S. equity markets in Rule 605 (formerly known as 11ac1-5) to release summary statistics about execution costs on a regular basis (see U.S. Securities and Exchange Commission (2000)).

Panel A of Table 5 examines the permanent price impact in subsamples formed by sorting on the average bid-ask spread of stocks into four quartiles. The spread is a commonly used measure of adverse selection or the extent of information asymmetry in the environment of the stock. We observe that in each quartile, the permanent price impact of orders coming through BG3 brokers, which we associate with domestic institutions, is larger than that of orders coming through BG1 brokers, which we associate with households. Since our working hypothesis is that domestic institutions are more likely to be informed, it appears that the market makes this inference and adjusts prices accordingly. It is interesting to note that the permanent price impact of orders that come from brokers that predominantly serve foreigners is generally greater than the permanent price impact of orders originating from BG1 brokers, but somewhat smaller in magnitude than that of orders originating from brokers that predominantly serve domestic institutions.

Since bid-ask spreads are used in Table 5 as a measure of the severity of the adverse selection problem, we would expect the permanent price impact of orders to increase as we move across the columns in the table. However, if the market does not consider household orders to be informative, orders coming from BG1 brokers should not exhibit the increasing pattern, while orders coming from BG3 brokers should exhibit it. This is exactly the picture we observe: The permanent price impact of BG1 orders is the same in Quartile 4 (henceforth, Q4) as in Quartile 1 (henceforth, Q1), and hardly changes across the spread quartiles. In contrast, the permanent price impact of BG3 orders, which predominantly come from better-informed domestic institutions, increases from 0.031 in small-spread stocks to 0.282 in large-spread stocks (t-statistic 11.2). We also observe that the permanent price impact of BG2 orders (which predominantly come from foreigners) increases significantly across the quartiles. The difference-in-differences tests in the last few rows of the table show that the permanent price impact of orders coming from brokers that predominantly serve domestic institutions increases across the spread quartiles more than that of orders coming from the other two broker groups.

A similar picture emerges from Panel B of Table 5 that looks at the permanent price impact of trades across quartiles formed by sorting on the firms' market capitalization. Like bid-ask spread, market capitalization has been used as a proxy for information asymmetry, where larger firms are said to have less information asymmetry or less informed trading (relative to total trading in the stock). The permanent price impact of orders coming from BG1 brokers is about the same in Q2, Q3, and Q4, with only a minor difference between the smallest firms and the rest of the firms in the market. In contrast, the permanent price impact of orders coming from brokers that predominantly serve domestic institutions decreases monotonically as firm size increases.

The cross-sectional tests in both panels of the table deliver the same two important results. First, we observe that prices adjust more to marketable orders from brokers that are associated with better-informed investors. Second, we observe no difference in the permanent price impact across subsamples that represent increased degree of information asymmetry for orders coming from brokers that are associated with uninformed investors, but a significant increase for orders coming from brokers associated with informed investors.

The last test that we report in this table relies on market-generated events in all stocks to create two subsamples with differing probabilities of informed trading instead of using differences in information asymmetry across stocks. The test is motivated by the model in Frino, Johnson, and Zheng (2005). In a simple sequential trade framework, they formalize the intuition that since informed traders have a strong incentive to trade multiple times in the same direction before prices adjust, a second consecutive buy order submitted by the same trader creates a larger permanent price impact than a second consecutive buy order that comes from a different trader. If the mix of clients of each broker is relatively constant, or if there are frictions that prevent informed investors from trading through multiple brokers, observing a sequence of buys with the same broker ID should cause a larger price adjustment than observing such a sequence with a mix of several broker IDs.

To test this idea, we look at marketable orders that follow a sequence of consecutive marketable orders in the same direction.[28] We then put each of these marketable orders in one of two categories. The first category (Dominant ID Sequences) consists of marketable orders that share the same broker ID with at least half of the sequence that preceded it (e.g., at least three marketable buy orders out of a sequence of five came from the same broker ID as the current marketable order). The second category (Mixed Sequences) contains marketable orders that follow sequences of marketable orders in the same direction but where there is no dominant broker ID.

In Panel C of Table 5 we compare the permanent price impact of the marketable orders in these two categories. We find that the permanent price impact of orders in the Dominant ID Sequences category for brokers that predominantly serve domestic institutions is much larger than that of orders in the Mixed Sequences category (0.093 versus 0.064; with a t-statistic of 5.0). A similar result (though with a smaller magnitude) is also observed for foreigners. In contrast, when the sequences come from a broker that predominantly serves households (i.e., it belongs to BG1), we observe that the permanent price impact of the Mixed Sequences category is greater (i.e., less negative) than that of the Dominant ID Sequences category. This means that market participants do not believe that these orders come from informed investors. Our evidence is therefore consistent with the prediction of the Frino et al. model only for broker IDs that are associated with informed investors, which echoes the findings of the cross sectional tests.[29]

4.2. Permanent Price Impact Regressions

Our findings in Table 5 could still potentially be explained by inferences made by market participants from different attributes of the orders that are unrelated to broker ID. For example, domestic institutions could be using larger orders or trading when the limit order book is thinner.

---

[28] We use all sequences of up to five consecutive marketable orders in the same direction that occurred within one hour.

[29] Frino et al. (2005) also carry out empirical work using trade sequences from the Australian Stock Exchange to test the predictions of their model. They find that a second transaction in the same direction has a greater impact on prices when the first transaction came from the same broker rather than another broker.

We therefore provide a more structured analysis of the permanent price impact that enables us to condition on order and market environment attributes in addition to broker ID. This is where we truly try to hold everything else equal and see whether otherwise-identical trades have larger (smaller) price impacts when they come through a broker that is usually associated with informed (uninformed) investors.

We estimate the following pooled regressions with stock fixed effects where the dependent variable (ppi) is the permanent price impact of trades:

$$\begin{aligned}\text{ppi}_{i,t} = {} & b_1\text{BG2}_{i,t} + b_2\text{BG3}_{i,t} + b_3\text{Volume}_{i,t} + b_4\text{SgnReturn}_{i,t} + b_5\text{Volatility}_{i,t} + \\ & b_6\text{TradeSize}_{i,t} + b_7\text{TradeSize}^2_{i,t} + b_8\text{SameSideDepth}_{i,t} + b_9\text{OtherSideDepth}_{i,t} + \\ & b_{10}\text{FirstTrade}_{i,t} + b_{11}\text{Duration}_{i,t} + b_{12}\text{Duration}^2_{i,t} + b_{13}\text{Spread}_{i,t} + b_{14}\text{Spread}^2_{i,t} + \\ & b_{15}\text{MarketTrade}_{i,t} + b_{16}\text{LargeTrade}_{i,t} + a_i\text{StockDummies}_{i,t} + \varepsilon_{i,t}\end{aligned} \tag{1}$$

The indices (i,t) refer to trade t in stock i. *BG2* and *BG3* are the dummy variables for the broker groups associated with foreigners and domestic institutions, respectively. As in the investor identity probit regressions of Section 3, the other variables belong to three categories. The first category describes recent market activity and includes volume, signed return, and volatility in the five minutes prior to the trade, as well as log duration and squared log duration.[30] These variables are included to control for the possibility that certain investors (e.g., institutions) are more active when there is more pricing-relevant information in the market. The second category describes the state of the limit order book and includes depth at the best bid and offer (same-side and other-side depth relative to the direction of the marketable order), as well as log bid-ask spread and squared log spread. The third category consists of measures of the size of the marketable order: (i) trade size, (ii) trade size squared, (iii) Market Trade dummy (which is equal to one if the size of the marketable order is equal to the number of shares available at the best

[30] The volume, signed return, and volatility are transformed for use in the permanent price impact regressions. The volume (defined as the number of shares traded in the previous five minutes divided by the number of shares outstanding) is multiplied by 10,000 for ease of economic interpretation and is winsorized at the 0.1% level. Both signed return and volatility are multiplied by 100 to keep the units of these variables the same as the units of the dependent variable. FirstTrade is a dummy variable that is set to one for the first trade of the day where we are unable to compute duration. The Duration variable is set to zero in these cases, and therefore the dummy variable FirstTrade shifts the intercept appropriately.

price in the book), and (iv) Large Trade dummy (which is equal to one if the size of the marketable order is larger than the number of shares available at the best price in the book).

The first column of Table 6 shows the results of this analysis for the entire sample. Note that orders coming from brokers that are associated with households serve as the (unreported) stock-specific intercept, and therefore the coefficients on BG2 and BG3 should be interpreted as the difference between the permanent price impact of orders coming from these brokers and the permanent price impact of BG1 orders. The coefficient on BG3 is positive and highly significant, consistent with a market inference that orders coming through brokers associated with domestic institutions are more likely to reflect private information. The coefficient on BG2 is positive but smaller than that on BG3, indicating that the market believes orders coming through brokers associated with foreigners could also be informative, but somewhat less than those coming from brokers that serve domestic institutions.

The table also reports separate regressions for market capitalization and bid-ask spread quartiles. We note that the coefficient on BG3 increases across bid-ask spread quartiles and decreases across market capitalization quartiles, in line with our unconditional results in Table 5. Hence, the same conclusions remain even after we control for various order and market environment attributes.[31]

4.3. Robustness

We performed several robustness tests to ensure that our results are not driven by the specific choices we make concerning the definition of the permanent price impact, the variables in the regressions, or the estimation methodology. A possible objection to defining the permanent price impact using an arbitrary length of time (i.e., five minutes) is that the amount of trading differs significantly in the cross-section of stocks. As Table 1 demonstrates, the average trading volume in the quartile of the largest stocks is over forty times the average trading volume in the quartile

[31] The coefficient on BG2 increases across the bid-ask spread quartiles from Q1 to Q3, but is small and statistically insignificant in Q4. We found that only a small number of orders are coming from BG2 brokers in this quartile, and the inherent noise in the data coupled with the small number of observations cause the estimate of this coefficient in Q4 to be unreliable.

of the smallest stocks. We therefore used another definition of permanent price impact whereby the time interval between the prevailing and the post-trade midquotes depends on the average activity level (in terms of number of trades) of the stocks.[32] The conclusions we draw from the empirical tests with this modified definition are similar to those using the regular definition of the permanent price impact presented in the tables. We also carried out our analysis with price impact measures computed from bid prices only or from ask prices only rather than using the midquotes (e.g., comparing the prevailing bid price to the bid price five minutes after the trade). The results, using either the five-minute rule or the definition of price impact with varying interval lengths, were similar to those presented in the tables.

One could raise an objection to the form of the regressions because the permanent price impact and other order attributes are determined together in equilibrium. While we are sympathetic to the equilibrium view, these regressions do not suffer from a simultaneity problem. The variables that reflect market activity and the state of the limit order book are clearly measured prior to the realization of the permanent price impact. While order size and duration could represent optimal choices given past values of the price impact, these variables could not, strictly speaking, be a function of the price impact of the marketable order because the price impact had not yet occurred when the choices on order size and timing were made. Therefore, the regressors are contemporaneously uncorrelated with the error, and OLS is consistent and retains its desirable asymptotic properties.

Nonetheless, we carried out two tests to examine the robustness of our results to this potential criticism. First, we ran the regressions without the independent variables that are functions of trade size or duration and our conclusions were unchanged. Second, we ran various

[32] More specifically, we divide the stocks in the sample into three groups based on natural cutoffs for the number of trades. We use a one-minute interval (from the prevailing midquote to the post-trade midquote) for the group of the most active stocks. The average daily number of trades (or marketable orders) for the most-active stock group is 184.40, while the average daily number of trades for the second group is 38.82. Since the first group has 4.75 times the number of trades of the second group, we use 4.75 minutes as the time interval for the second group. Using a similar calculation we arrive at a time interval of 17.43 minutes between the prevailing and the post-trade midquote for the least actively traded group of stocks.

vector autoregression specifications to evaluate the robustness of the results we obtain from the single-equation permanent price impact regressions. Our focus on the relationship between the permanent price impact and broker identity necessitated VAR and VARX specifications where BG2 and BG3 were also included as endogenous variables with their own equations. We performed the vector autoregressions on a stock-by-stock basis and examined the resulting coefficients, Granger-causality tests, and impulse response functions. We found that the lagged price impact did not play a significant role in the broker ID equations, validating the single-equation approach. The impulse response functions demonstrated that most of the price adjustment occurs very quickly (within a few trades).

As another robustness check we estimated the permanent price impact regressions separately for marketable orders of varying sizes. As in Figure 2, we created four trade-size quartiles based on stock-specific breakpoints so that each stock contributes observations to all quartiles. Figure 3 plots the coefficients on BG2 and BG3 from these regressions, and we observe an interesting pattern: When an order comes from brokers associated with domestic institutions, the market infers that the information content of a medium-size order is greater than that of a large-size order.[33] A similar pattern arises when the inference is about marketable orders coming from foreigners. This result is consistent with the "stealth trading" hypothesis of Barclay and Warner (1993), whereby informed investors choose to split their desired (large) position change into multiple medium-size trades in order to hide among the uninformed investors. While in principle lack of anonymity could either increase the need to use order splitting (because broker ID makes it more difficult to hide) or decrease it (because it is futile), the evidence we document is consistent with the former.

Lastly, we added dummy variables for the first and last hours in the day to control for deterministic patterns in price impact and the possibility of a richer information environment at the beginning of the day. We also used past one-hour return in an attempt to control for periods

[33] While we plot only the coefficients on the dummy variables, these regressions also include all the explanatory variables and stock fixed effects as in Table 5.

in which information events are more likely (similar to the past five-minute return in the tables), and the results of our analysis were unaffected by the inclusions of these additional controls.

4.4. Interactions

Our findings in the previous section demonstrate that market participants make pricing-relevant inferences from broker ID over and above the inference that can be made from other order characteristics and market conditions. However, investors could be utilizing broker ID in conjunction with other attributes of the order in a more sophisticated manner that refines the signal about investor types. In other words, an attribute of the order or the trading environment could mean different things depending on whether or not the market believes the order comes from an institution as opposed to an individual, or alternatively, other attributes could help market participants reduce misclassification errors that are inherent in using broker identity to represent investor types.

Consider the following example. We know from Table 2 that 72.3% of the marketable orders coming through BG1 brokers are indeed household orders, while 20.8% come from domestic institutions. Using the broker ID of BG1 brokers as a signal for household orders is therefore subject to an error. Table 7 demonstrates how using trade size could reduce those inference errors. As in Table 2, we look at the fractions of trades that come from each investor type through each broker group, but here we do it separately for four trade-size quartiles. The table shows that 81.7% of the smaller BG1 orders (Q1) are household orders. However, only 40% of the larger BG1 orders (Q4) originate from households, while 43.3% originate from domestic institutions and 16.8% from foreigners. In fact, for this order-size category, the likelihood of observing a more informed investor trading via a broker that belongs to BG1 is greater than the likelihood of observing a household, which is also consistent with the results of the probit regressions presented in Figure 2.

We therefore look for evidence concerning this potentially sophisticated inference by including the interactions between broker ID and other order and market environment attributes in the price impact regressions. Since regressions with dozens of interactions lend themselves to a rather messy presentation, we use a couple of simpler specifications. In Panel A of Table 8 we run the price impact regressions separately for one broker group dummy (and its interactions) at a time.[34] We estimate the following pooled regression with stock fixed effects and with the same explanatory variables as in Table 6:

$$
\begin{aligned}
\mathrm{ppi}_{i,t} = {} & b_1 \mathrm{Volume}_{i,t} + b_2 \mathrm{SgnReturn}_{i,t} + b_3 \mathrm{Volatility}_{i,t} + b_4 \mathrm{TradeSize}_{i,t} + b_5 \mathrm{TradeSize}^2_{i,t} + \\
& b_6 \mathrm{SameSideDepth}_{i,t} + b_7 \mathrm{OtherSideDepth}_{i,t} + b_8 \mathrm{FirstTrade}_{i,t} + b_9 \mathrm{Duration}_{i,t} + \\
& b_{10} \mathrm{Duration}^2_{i,t} + b_{11} \mathrm{Spread}_{i,t} + b_{12} \mathrm{Spread}^2_{i,t} + b_{13} \mathrm{MarketTrade}_{i,t} + b_{14} \mathrm{LargeTrade}_{i,t} + \\
& \mathrm{BG}_{i,t} \times \begin{pmatrix} c_0 + c_1 \mathrm{Volume}_{i,t} + c_2 \mathrm{SgnReturn}_{i,t} + c_3 \mathrm{Volatility}_{i,t} + c_4 \mathrm{TradeSize}_{i,t} + c_5 \mathrm{TradeSize}^2_{i,t} + \\ c_6 \mathrm{SameSideDepth}_{i,t} + c_7 \mathrm{OtherSideDepth}_{i,t} + c_8 \mathrm{FirstTrade}_{i,t} + c_9 \mathrm{Duration}_{i,t} + \\ c_{10} \mathrm{Duration}^2_{i,t} + c_{11} \mathrm{Spread}_{i,t} + c_{12} \mathrm{Spread}^2_{i,t} + c_{13} \mathrm{MarketTrade}_{i,t} + c_{14} \mathrm{LargeTrade}_{i,t} \end{pmatrix} \\
& + a_i \mathrm{StockDummies}_{i,t} + \varepsilon_{i,t}
\end{aligned}
\tag{2}
$$

While the table provides the coefficients for all variables in the regression (except for the fixed effects), we focus our attention on the last three columns that present the coefficients on the interactions. For example, in the BG1 regression we find a positive and highly significant coefficient on the interaction between trade size and the broker ID dummy variable, which means that larger marketable orders have larger permanent price impacts when they come from brokers that predominantly serve households. This could reflect two effects: (i) households who use larger marketable orders are more likely to have useful information (Easley and O’Hara (1987)), and (ii) the correct inference from a large marketable order that comes through a BG1 broker is that the order originated from a domestic institution rather than from a household. On the other hand, the coefficients on the interaction of trade size with BG2 and BG3 are negative, which means that if the order is more likely to come from an institution, a large order is thought of as less informed. This is consistent with the “stealth trading” result of Barclay and Warner (1993) that was noted earlier and whereby most information is brought into the market by medium-size (rather than large-size) trades.

[34] We note that our conclusions are the same when we run a different specification that includes multiple dummy variables and their interactions in one model.

Other interactions also demonstrate that the inference from the arrival of a marketable order is more complex and involves broker ID together with order or market attributes. We know from Table 6 that the unconditional effect of duration on the permanent price impact is negative: The price impact is smaller when the duration between trades is longer. This result is consistent with the theoretical model of Easley and O'Hara (1992) and the empirical findings in Dufour and Engle (2000), whereby an increase in the pace of trading indicates a greater likelihood that an information event has occurred and therefore an increase in the probability of informed trade. The interaction between broker ID and duration in Panel A of Table 8 shows that when orders come from brokers that predominantly serve domestic institutions (and foreigners), this effect is further enhanced: The interaction is negative and significant. However, when market participants see orders from brokers that predominantly serve households, they infer that this order flow is probably uninformed. Hence, shorter duration (i.e., an increase in the pace of trading) reflects more "noise" trading, and the permanent price impact decreases.

Table 6 also shows that the unconditional effect of the bid-ask spread is positive: Trades tend to have larger permanent price impacts when the prevailing spread is wide. This is not surprising as the spread ought to reflect the extent of adverse selection in the environment (see, for example, Glosten and Milgrom (1985) and Easley and O'Hara (1987)). The coefficient on the interaction of BG3 and the spread in Panel A of Table 8 is also positive as expected. However, we observe that the coefficient on the interaction of BG1 and the spread is negative: Marketable orders that come through brokers that are associated with households when the spread is wider tend to have smaller permanent price impacts. In other words, the inference from a BG1 order is that the investor behind it is uninformed and therefore the order should not cause a permanent price impact. The reduction in the expected permanent price impact due to the inference that the order constitutes liquidity (or "noise") trading is greater when the extent of adverse selection (as reflected in the spread) is larger.

Our findings in Panel A of Table 8 concerning the interaction of broker identity and other attributes of the order highlight the delicate and remarkable inference work performed by the market. Anything that deviates from the usual behavior of an investor type is treated very differently by the market and results in a different permanent price impact pattern. In particular, we observe that the manner in which attributes of the order and the trading environment affect how prices evolve critically depends on the inference the market makes from broker ID concerning the likelihood that the investor behind the trade is informed.

The last test in this section attempts to determine whether it is ultimately the investor type that matters or only the identity of the broker that seems to affect prices for some unknown reason. We examine this question in Panel B of Table 8 by replacing the broker group dummy variables used in the regressions with a continuous variable that represents historical trading by investor types. For example, in the households regression we compute for each broker in each stock the fraction of trades that come from households. Hence, the number of distinct values this variable takes in the pooled regression is the number of brokerage firms times the number of stocks, and a higher value of this variable means that there is a greater likelihood that a trade was initiated by a household rather than by a domestic institution or a foreigner. We construct this variable in an analogous fashion for the foreigners and domestic institutions regressions.

The results of the analysis with the continuous investor type variables turn out to be very similar to those using the broker group dummy variables: The coefficients on the interactions have the same signs and are rather similar in terms of statistical significance. While there is no doubt that the customer base variable is related to the broker group dummy variable because of the manner in which we created the broker groups, the two do not contain exactly the same information. First, the continuous variable provides broker-level information (rather than broker-group-level information) about investor types. Second, and most importantly, unlike broker ID the customer base variable is not directly observable to the market. That the two yield such similar results contributes to our confidence in interpreting the broker ID results as driven by the inference market participants make about investor types.

5. Conclusions

As stock exchanges and trading venues around the world tinker with the design of their execution mechanisms and the rules that govern trading, the degree of anonymity in the market becomes an important feature to consider. As researchers, we want to determine whether signals about the identity of investors indeed make a difference with respect to market prices. This question is especially important because we would like to understand why event studies document improvement in market liquidity when exchanges adopt a more anonymous market structure. Our findings suggest an answer to this question: Prices adjust less efficiently to order flow information in completely anonymous markets than in markets that feature an intermediate level of anonymity.

While we document that investors use multiple brokers in an apparent attempt to hide their trading motives, our study demonstrates that broker identity information can nonetheless be used as a powerful signal to help uncover attributes of investors who send orders to the market. The implication of this finding is that there must be frictions in the economic environment that prevent investors from utilizing multiple brokers to such an extent that the information content of broker ID is eliminated.

There are several important practices in the financial services industry that could be the source of these frictions. In particular, customer bases may vary across brokers because brokerage firms differ with respect to the levels of service they offer (e.g., the services offered by a full-service broker versus those offered by a discount broker) and investors have heterogeneous preferences over the breadth of available services. For example, one way in which soft-dollar commissions create a drag on institutional investor profitability is that in order to obtain research from a full-service broker, the institution needs to direct a significant amount of order flow to the broker and give up the opportunity to hide in the order flows of other brokers. Similarly, institutions that utilize special order-management systems developed and maintained by particular brokers face hurdles that make it more difficult to trade with many brokers simultaneously.

In principle, any variation in brokerage fees might lead certain types of investors to favor broker A over broker B. For example, quantity discounts that lower the fees to heavy traders when they send many orders to the same broker rather than sending fewer orders through multiple brokers operate to counter the benefit informed investors might enjoy by utilizing a mixed strategy over brokers. Furthermore, an investor may benefit from channeling substantial business to a broker because the cost of executing a trade partially depends on the effort of the broker, and brokers may be more inclined to exert effort for bigger clients. If informed investors behave optimally, it should be the case that their losses from not utilizing multiple brokers to hide are equal to or less than the benefits obtained from the aforementioned services or discounts.

Goldstein, Irvine, Kandel, and Wiener (2009) use data on commissions and volume allocations of institutional investors to argue that the brokerage industry is characterized by a high degree of bundling that prevents order flow from migrating to the broker with the lowest commissions. We use information on broker identity and prices to demonstrate that informed investors do not migrate to the brokers that would afford them the lowest price impact of trading, which further testifies to the significance of these frictions.

Trading practices are currently undergoing dramatic changes driven by technological innovations (see, for example, Hasbrouck and Saar (2010)). Agency algorithmic trading—the use of computer algorithms to achieve better trade execution—is gaining wide acceptance among institutional investors in financial markets. Full-service brokers develop such systems for use by their institutional clients, which raises the sophistication of the average institutional order flow and enables informed investors to "hide" better. This trend would most likely cause the difference between households and institutions that we document to persist and become even more pronounced with time. Still, computer algorithms can also be very efficient in "uncovering" informed order flow by translating the broker ID signal in conjunction with other attributes of the environment into strategies that proprietary high-frequency traders can utilize profitably. At some point, such efficient utilization of broker ID signals by algorithms will lower the profitability of informed investors enough to trigger increased use of mixed strategies over multiple brokers even in the face of frictions. The degree of anonymity of the trading mechanism as a market design feature will no doubt continue to be a major determinant of both the strategic interaction between investors and the evolution of prices in years to come.

6. References

Aspris, Angelo, Alex Frino, Dionigi Gerace, and Andrew Lepone, 2008, Limit order book, anonymity, and market liquidity: Evidence from the Sydney Futures Exchange, *Accounting and Finance* 48, 561-573.

Admati, Anat R., and Paul Pfleiderer, 1991, Sunshine trading and financial market equilibrium, *Review of Financial Studies* 4, 443-481.

Barclay, Michael J., and Jerold B. Warner, 1993, Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics* 34, 281-305.

Battalio, Robert, Andrew Ellul, and Robert Jennings, 2007, Reputation effects in trading on the New York Stock Exchange, *Journal of Finance* 62, 1243-1271.

Battalio, Robert, and Craig W. Holden, 2001, A simple model of payment for order flow, internalization, and total trading cost, *Journal of Financial Markets* 4, 33-71.

Benveniste, Lawrence M., Alan J. Marcus, and William J. Wilhelm, 1992, What's special about the specialist? *Journal of Financial Economics* 32, 61-86.

Boulatov, Alex, and Thomas J. George, 2008, Securities trading when liquidity providers are informed, Working paper, University of Houston.

Chan, Louis, K. C., and Joseph Lakonishok, 1995, The behavior of stock prices around institutional trades, Journal of Finance 50, 1147-1174.

Choe, Hyuk., Bong-Chan Kho, and Rene M. Stulz, 2005, Do domestic investors have an edge? The trading experience of foreign investors in Korea, *Review of Financial Studies* 18, 795-829.

Comerton-Forde, Carole, Alex Frino, and Vito Mollica, 2005, The impact of limit order anonymity on liquidity: Evidence from Paris, Tokyo, and Korea, *Journal of Economics and Business* 57, 528-540.

Dufour, Alfonso, and Robert F. Engle, 2000, Time and the price impact of a trade, *Journal of Finance* 55, 2467-2498.

Easley, David, and Maureen O’Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69-90.

Easley, David, and Maureen O’Hara, 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577-605.

Foster, Margaret M., and Thomas J. George, 1992, Anonymity in securities markets, *Journal of Financial Intermediation* 2, 168-206.

Foucault, Thierry, Sophie Moinas, and Erik Theissen, 2007, Does anonymity matter in electronic limit order markets? *Review of Financial Studies* 20, 1707-1747.

Frino, Alex, David Johnstone, and Hui Zheng, 2005, The information content of trader identification, Working paper, University of Sydney.

Garfinkel, Jon A., and M. Nimalendran, 2003, Market structure and trader anonymity: An analysis of insider trading, *Journal of Financial and Quantitative Analysis* 38, 591-610.

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.

Goldstein, Michael A., Paul Irvine, Eugene Kandel, and Zvi Wiener, 2009, Brokerage commissions and institutional trading patterns, *Review of Financial Studies* 22, 5175-5212.

Grinblatt, Mark, and Matti Keloharju, 2000, The investment behavior and performance of various investor types: A study of Finland’s unique data set, *Journal of Financial Economics* 55, 43-67.

Hasbrouck, Joel, 2009, *Empirical Market Microstructure* (Oxford University Press, New York, NY).

Hasbrouck, Joel, and Gideon Saar, 2010, Low-Latency Trading, Working paper, the Johnson School, Cornell University.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1336.

Madhavan, Ananth, 1996, Security prices and market transparency, *Journal of Financial Intermediation* 5, 255-283.

Maher, Owen, Peter L. Swan, and P. Joakim Westerholm, 2008, Twilight falls on the limit order book: Endogeneity and the demise of broker identity, Working paper, University of New South Wales.

Majois, Christophe, 2009, Liquidity in limit-order markets: Revisiting the switch to anonymity on Euronext Paris, Working paper, Louvain School of Management.

Rindi, Barbara, 2008, Informed traders as liquidity providers: Anonymity, liquidity and price formation, *Review of Finance* 12, 497-532.

Röell, Ailsa, 1990, Dual-capacity trading and the quality of the market, *Journal of Financial Intermediation* 1, 105-124.

Shukla, R. K., and G. B. van Inwegen, 1995, Do domestics perform better than foreigners? An analysis of UK and US mutual fund managers, *Journal of Economics and Business* 47, 241-254.

U.S. Securities and Exchange Commission, 2000, Final Rule: Disclosure of Order Execution and Routing Practices, available online at http://www.sec.gov/rules/final/34-43590.htm.

Table 1
Sample Summary Statistics

Our sample period starts on July 10, 2000 and ends on October 23, 2001. Our universe of securities includes all stocks that were listed on the Helsinki Stock Exchange during the sample period (i.e., we exclude warrants, employee stock options, and temporary share classes that are traded on the exchange). We compute the number of trades per listed firm, and retain a listed firm in the sample if it has at least five trades per day on average during our sample period. If a listed firm has multiple share classes, we aggregate across share classes for the purpose of computing the daily average number of trades. This trading activity screen results in 87 listed firms that constitute the sample. The following variables are calculated for each firm over the sample period: MktCap is the average market capitalization of the firms (in million Euros), AvgVol is the average daily Euro volume (in thousands), StdRet is the standard deviation of daily returns, and AvgSprd (in %) is the average relative spread (the Euro spread divided by the midpoint between the best bid and ask prices). The table presents the cross-sectional medians of these variables for the entire sample and separately for quartiles sorted by market capitalization, average daily volume, return standard deviation, and average relative bid-ask spread.

		Number of Firms	**MktCap** (million Euros)	**AvgVol** (1,000 Euros)	**StdRet** (in %)	**AvgSprd** (in %)
Entire Sample		87	355.93	168.79	3.94%	1.54%
MktCap Groups	Q1 (Low)	21	48.53	31.73	5.01%	2.84%
	Q2	22	178.24	126.95	4.45%	1.92%
	Q3	21	523.83	253.16	2.49%	1.37%
	Q4 (High)	23	1888.80	1430.00	2.61%	0.77%
Volume Groups	Q1 (Low)	21	52.50	27.92	4.26%	2.89%
	Q2	22	255.99	96.97	3.82%	2.18%
	Q3	21	523.83	317.95	3.00%	1.27%
	Q4 (High)	23	1514.65	1816.77	2.76%	0.71%
StdRet Groups	Q1 (Low)	21	709.64	285.11	2.16%	1.12%
	Q2	22	581.11	148.44	2.81%	1.66%
	Q3	21	142.85	90.86	4.44%	2.38%
	Q4 (High)	23	143.99	152.46	5.84%	2.01%
Spread Groups	Q1 (Low)	21	1514.65	1850.68	2.76%	0.67%
	Q2	22	522.54	395.80	2.51%	1.24%
	Q3	21	236.85	90.86	3.00%	2.01%
	Q4 (High)	23	60.10	31.73	4.78%	3.20%

Table 2
Broker Clienteles

This table shows how much trading of each investor type goes through the three broker groups. We consolidate all investors into three types: (i) all domestic households, (ii) foreign investors who trade via nominee accounts, and (iii) domestic institutional investors (i.e., finance and insurance institutions, non-financial corporations, nonprofits, and government). We use a simple rule to categorize brokerage firms: A broker belongs to a Broker Group that is associated with a certain investor type (BG1 with households, BG2 with foreigners, and BG3 with domestic institutions) if more than 50 percent of the trades that the broker executes involve this investor type. This rule is applied on a stock-by-stock basis because some brokers could see the majority of their orders coming from one investor type (e.g., domestic institutions) in large and active stocks, while having most of their trading in thinly traded stocks come from another investor type (e.g., households). The cells of the table report the fractions of trades (and volume) in each broker group that come from the actual investor types. The analysis in the columns labeled "Investor Categories (EW)" first computes the fraction separately for each firm, and then averages across firms. The analysis in the columns "Investor Categories (UW)" simply sums the total number of trades (or volume) in each investor category for each broker group, and computes the fractions from these aggregate numbers.

		Investor Category (EW)			**Investor Category (UW)**		
	Broker Group	Households	Foreigners	Domestic Institutions	Households	Foreigners	Domestic Institutions
Number of Trades	BG1	72.2%	5.6%	22.2%	72.3%	6.9%	20.8%
	BG2	4.9%	77.6%	17.5%	3.7%	79.1%	17.2%
	BG3	19.0%	13.9%	66.0%	11.5%	16.9%	71.6%
Trading Volume	BG1	55.6%	10.2%	34.2%	50.0%	16.6%	33.3%
	BG2	2.3%	80.0%	17.7%	0.7%	82.5%	16.8%
	BG3	11.1%	15.5%	72.3%	2.1%	17.9%	79.9%

Table 3
Use of Multiple Brokers by Investors

This table provides evidence on the use of multiple brokers by households and domestic institutions. Each line of the table uses a different definition of the term "multi-broker user." In the first line, a multi-broker user is defined as an investor who uses more than one brokerage firm during a single day on at least 25% of the days in which the investor is active (i.e., sending orders to the exchange). In the second line, a multi-broker user is defined as an investor who uses more than one brokerage firm to trade the same stock during the same week in at least 25% of the weeks in which the investor is active. In the third line, a multi-broker user is defined as an investor who uses brokers from at least two different broker groups to trade the same stock during the same week in at least 25% of the weeks in which the investor is active. In the fourth line, a multi-broker user is defined as an investor who uses brokers from at least two different broker groups to trade the same stock during a single day on at least 25% of the days in which the investor is active. We use a simple rule to categorize brokerage firms into broker groups: A broker belongs to a broker group that is associated with a certain investor type if more than 50 percent of the trades that the broker executes involve this investor type. This rule is applied on a stock-by-stock basis because some brokers could see the majority of their orders coming from one investor type (e.g., domestic institutions) in large and active stocks, while having most of their trading in thinly traded stocks come from another investor type (e.g., households). The cells of the table report for each investor type (i) the percentage of multi-broker users among the investors who belong to that type, and (ii) the fraction of trades of the investor type that originate from multi-broker users.

Definition of Multi-Broker User	**Households**		**Domestic Institutions**	
	Investors	Trades	Investors	Trades
Multiple Brokers, Any Stock, Same Day	9.1%	20.7%	11.5%	81.0%
Multiple Brokers, Same Stock, Same Week	8.1%	17.3%	12.0%	68.1%
Multiple Broker Groups, Same Stock, Same Week	1.6%	6.6%	3.8%	55.1%
Multiple Broker Groups, Same Stock, Same Day	1.1%	3.2%	1.7%	42.2%

Table 4
Investor Identity Probit Models

This table reports the results of probit regressions that examine whether broker identity provides useful information about investor types over and above the information in order and market environment attributes. Each investor who submits a marketable limit order that triggers a transaction is classified into one of three categories: households, foreigners, and domestic institutions. In the columns labeled Households, the dependent variable in the probit regression is set to one if a household submits the marketable limit order and to zero if either a foreigner or a domestic institution submits the marketable limit order. Similarly, the dependent variable in the Foreigners (Domestic Institutions) columns is set to one if a foreigner (domestic institutions) submits the marketable limit order and to zero if the order was submitted by a different type of investor. Explanatory variables consist of broker group dummies (BG1, BG2, and BG3) and a set of controls. Control variables include recent activity (volume, signed return, and volatility in the previous five minutes; log duration from the previous trade and log duration squared; and a dummy variable that is set to one for the first trade of the day where we are unable to compute duration), prevailing state of the limit order book (same-side and other-side depth at the best price; log prevailing spread and squared log spread), and trade size (log and squared log of trade size; a dummy variable for trades that are equal to the displayed depth at the best price, and a dummy variable for marketable orders that are larger than the displayed depth). The probit regressions are estimated using observations pooled from all stocks and include stock fixed effects that are omitted from the table for ease of presentation. We use "**" to indicate significance at the 1% level and "*" to indicate significance at the 5% level (both against a two-sided alternative).

		Households		Foreigners		Domestic Institutions	
BG1	Coef.	0.2123**	0.3038**	-1.3568**	-1.0345**	-0.4896**	-0.2396**
	t-stat.	(21.29)	(29.84)	(-120.07)	(-85.71)	(-54.41)	(-25.02)
BG2	Coef.	-2.1319**	-1.7174**	0.8487**	1.0291**	-0.7838**	-0.7434**
	t-stat.	(-204.92)	(-160.43)	(74.88)	(85.70)	(-85.77)	(-77.57)
BG3	Coef.	-1.5413**	-1.0784**	-0.9872**	-0.8018**	0.8795**	0.9514**
	t-stat.	(-143.85)	(-97.96)	(-85.31)	(-65.58)	(93.79)	(96.85)
Volume	Coef.		-0.0572		2.9140**		11.0215**
	t-stat.		(-0.002)		(3.05)		(9.36)
SgnReturn	Coef.		-0.5989**		1.6888**		-0.5638**
	t-stat.		(-3.94)		(13.64)		(-5.26)
Volatility	Coef.		0.6566**		-4.5562**		2.9784**
	t-stat.		(5.13)		(-40.69)		(31.46)
Trade Size	Coef.		0.0659**		0.4316**		0.6408**
	t-stat.		(7.03)		(86.54)		(128.81)
Trade Size Squared	Coef. / t-stat.		-0.0248** (-49.97)		-0.0153** (-67.57)		-0.0255** (-112.64)
SameSide BBO Depth	Coef. / t-stat.		-0.0007 (-0.861)		-0.0044** (-9.98)		0.0018 (1.78)
OtherSide BBO Depth	Coef. / t-stat.		-0.0000 (-0.138)		-0.00002 (-0.988)		0.00002 (1.08)
FirstTrade	Coef.		0.4048**		-0.1681**		-0.2156**
	t-stat.		(37.07)		(-13.40)		(-21.52)
Duration	Coef.		0.0646**		0.0619**		-0.0805**
	t-stat.		(30.28)		(39.60)		(-56.23)
Duration Squared	Coef. / t-stat.		-0.0039** (-15.33)		-0.0086** (-40.63)		0.0076** (41.00)
Spread	Coef.		-8.7348**		4.6477**		4.5111**
	t-stat.		(-34.13)		(16.50)		(20.83)
Spread Squared	Coef. / t-stat.		37.1261** (14.37)		-29.3577** (-8.93)		-13.1845** (-6.05)
Market Trade	Coef. / t-stat.		-0.5407** (-126.79)		0.0365** (10.86)		0.3036** (101.70)
Large Trade	Coef. / t-stat.		-0.0285** (-9.76)		-0.0108** (-5.04)		0.0437** (21.28)
Pseudo-R^2		0.3277	0.4395	0.2906	0.3045	0.1314	0.1577

Table 5
Permanent Price Impact

This table reports the permanent price impact of marketable limit orders coming from the three broker groups for several subsamples that differ by the extent of information asymmetry. The permanent price impact is defined as the signed log-change in the midpoint of the quote from an instant before the transaction to five minutes after the transaction. The price impact measure is standardized by subtracting the mean price impact in the same stock for marketable orders in the same direction (i.e., separate means are computed for buys and sells). For ease of interpretation, the permanent price impact measure is multiplied by 100, so a value of "1" indicates that the (standardized) log-change in the quote midpoint is 1%. In Panel A, the information asymmetry proxy is the average bid-ask spread of the stock over the sample period. We sort all stocks into four quartiles according to the spread measure and report the average permanent price impact in each quartile. In Panel B, the information asymmetry proxy is the average market capitalization of the firm over the sample period. In Panel C, we define two categories with different likelihood of informed trading by looking at whether the majority of marketable orders in a sequence of same-direction orders come from the same or different brokers. We report three types of t-tests. In the last column of the table, we test whether the permanent price impact of the same broker group differs between quartile 1 and quartile 4 of the information asymmetry category. In the fourth through sixth rows we test whether the permanent price impact in the same quartile differs between the different broker groups. In the seventh to ninth rows we present "difference-in-differences" tests to see whether the difference in the price impact between quartiles 1 and 4 for a certain broker group is different from that difference for another broker group. We also show the F-value for the ANOVA interaction of broker groups and the information asymmetry categories.

Panel A: Price Impact of Marketable Orders by Spread Quartiles

Broker Group	**Bid-Ask Spread Quartiles**				
	Q1(smallest)	Q2	Q3	Q4 (largest)	$\text{t-test}_{(Q4-Q1)}$
BG1 (households)	-0.069	-0.052	-0.070	-0.069	-0.1
BG2 (foreigners)	0.017	0.117	0.198	0.196	7.8
BG3 (domest. inst.)	0.031	0.134	0.218	0.282	11.2
$\text{t-test}_{(BG1-BG2)}$	-72.8	-33.0	-25.1	-10.8	
$\text{t-test}_{(BG1-BG3)}$	-59.7	-26.7	-24.6	-14.6	
$\text{t-test}_{(BG2-BG3)}$	-10.1	-2.1	-1.3	-2.7	
t-test	$BG1_{(Q4-Q1)}-BG2_{(Q4-Q1)} = -7.3$				
t-test	$BG1_{(Q4-Q1)}-BG3_{(Q4-Q1)} = -10.4$				
t-test	$BG2_{(Q4-Q1)}-BG3_{(Q4-Q1)} = -2.2$				
ANOVA	Broker Groups * Bid-Ask Spread Quartiles = 367.3				

Panel B: Price Impact of Marketable Orders by Market Capitalization Quartiles

Broker Group	**Market Capitalization Quartiles**				
	Q1(smallest)	Q2	Q3	Q4 (largest)	$\text{t-test}_{(Q4-Q1)}$
BG1 (households)	-0.056	-0.065	-0.064	-0.069	-2.8
BG2 (foreigners)	0.354	0.142	0.118	0.011	-21.7
BG3 (domest. inst.)	0.267	0.175	0.134	0.022	-14.7
$\text{t-test}_{(BG1-BG2)}$	-25.1	-37.5	-48.0	-62.6	
$\text{t-test}_{(BG1-BG3)}$	-18.8	-29.1	-31.8	-53.3	
$\text{t-test}_{(BG2-BG3)}$	3.8	-3.6	-2.4	-8.5	
t-test	$BG1_{(Q4-Q1)}-BG2_{(Q4-Q1)} = 20.1$				
t-test	$BG1_{(Q4-Q1)}-BG3_{(Q4-Q1)} = 13.5$				
t-test	$BG2_{(Q4-Q1)}-BG3_{(Q4-Q1)} = -4.3$				
ANOVA	Broker Groups * Market Cap. Quartiles = 594.3				

Panel C: Price Impact of Marketable Orders by Broker ID Sequence Categories

Broker Group	Categories: Dominant ID Sequences	Categories: Mixed Sequences	t-test$_{(Dom\text{-}Mix)}$
BG1 (households)	-0.066	-0.045	-6.4
BG2 (foreigners)	0.048	0.033	8.5
BG3 (domest. inst.)	0.093	0.064	5.0
t-test$_{(BG1\text{-}BG2)}$	-42.2	-33.9	
t-test$_{(BG1\text{-}BG3)}$	-40.9	-25.7	
t-test$_{(BG2\text{-}BG3)}$	-13.9	-8.0	
t-test	$BG1_{(Dom\text{-}Mix)}$-$BG2_{(Dom\text{-}Mix)}$ = -9.6		
t-test	$BG1_{(Dom\text{-}Mix)}$-$BG3_{(Dom\text{-}Mix)}$ = -7.6		
t-test	$BG2_{(Dom\text{-}Mix)}$-$BG3_{(Dom\text{-}Mix)}$ = -2.3		
ANOVA	Broker Groups * Sequence Categories = 74.9		

Table 6
Permanent Price Impact Regressions

This table presents regressions of the permanent price impact on broker group dummy variables and various controls. The regressions are presented separately for the entire sample and for market capitalization and bid-ask spread quartiles. The permanent price impact (ppi) is defined as the signed log-change in the midpoint of the quote from an instant before the transaction to five minutes after the transaction. The price impact measure is standardized by subtracting the mean price impact in the same stock for marketable orders in the same direction (i.e., separate means are computed for buys and sells). For ease of interpretation, the permanent price impact measure is multiplied by 100, so a value of "1" indicates that the (standardized) log-change in the quote midpoint is 1%. We estimate the following pooled regressions with stock fixed effects where the dependent variable is the permanent price impact of trades:

$$\begin{aligned}\text{ppi}_{i,t} = {} & b_1\text{BG2}_{i,t} + b_2\text{BG3}_{i,t} + b_3\text{Volume}_{i,t} + b_4\text{SgnReturn}_{i,t} + b_5\text{Volatility}_{i,t} + b_6\text{TradeSize}_{i,t} + b_7\text{TradeSize}^2_{i,t} + \\ & b_8\text{SameSideDepth}_{i,t} + b_9\text{OtherSideDepth}_{i,t} + b_{10}\text{FirstTrade}_{i,t} + b_{11}\text{Duration}_{i,t} + b_{12}\text{Duration}^2_{i,t} + b_{13}\text{Spread}_{i,t} + \\ & b_{14}\text{Spread}^2_{i,t} + b_{15}\text{MarketTrade}_{i,t} + b_{16}\text{LargeTrade}_{i,t} + a_i\text{StockDummies}_{i,t} + \varepsilon_{i,t}\end{aligned}$$

The indices (i,t) refer to trade t in stock i. *BG2* and *BG3* are the dummy variables for the broker groups associated with foreigners and domestic institutions, respectively. The other variables belong to three categories. The first category describes recent market activity and includes volume, signed return, and volatility in the five minutes prior to the trade, as well as log duration and squared log duration. FirstTrade is a dummy variable that is set to one for the first trade of the day where we are unable to compute duration. The Duration variable is set to zero in these cases, and therefore the dummy variable FirstTrade shifts the intercept appropriately. The second category describes the state of the limit order book and includes depth at the best bid and offer (same-side and other-side depth relative to the direction of the marketable order), as well as log bid-ask spread and squared log spread. The third category consists of measures of the size of the marketable order: (i) trade size, (ii) trade size squared, (iii) Market Trade dummy (which is equal to one if the size of the marketable order is equal to the number of shares available at the best price in the book), and (iv) Large Trade dummy (which is equal to one if the size of the marketable order is larger than the number of shares available at the best price in the book). We use "**" to indicate significance at the 1% level and "*" to indicate significance at the 5% level (both against a two-sided alternative).

		Full Sample	Market Capitalization Q1(small)	Q2	Q3	Q4(large)	Bid-Ask Spread Q1(small)	Q2	Q3	Q4(large)
BG2	Coef.	0.040**	0.106**	0.055**	0.049**	0.036**	0.044**	0.051**	0.085**	0.011
	t-stat.	(23.08)	(5.60)	(7.98)	(9.81)	(21.61)	(26.85)	(7.82)	(6.76)	(0.40)
BG3	Coef.	0.067**	0.126**	0.130**	0.109**	0.049**	0.060**	0.085**	0.159**	0.191**
	t-stat.	(28.91)	(6.47)	(14.09)	(13.81)	(23.29)	(28.19)	(10.07)	(11.76)	(7.17)
Volume	Coef.	0.162**	0.250**	0.122**	0.200**	0.100**	0.126**	0.111**	0.138**	0.168**
	t-stat.	(50.26)	(28.51)	(15.19)	(23.89)	(16.63)	(26.55)	(14.99)	(16.33)	(13.86)
SgnReturn	Coef.	-0.060**	-0.094**	-0.038**	-0.008	-0.072**	-0.055**	-0.035**	-0.106**	-0.104**
	t-stat.	(-25.01)	(-9.85)	(-5.87)	(-1.22)	(-27.00)	(-21.28)	(-5.71)	(-9.62)	(-7.73)
Volatility	Coef.	0.024**	0.050**	0.022**	0.024**	0.014**	0.016**	0.030**	0.079**	0.028
	t-stat.	(9.29)	(5.55)	(3.45)	(3.76)	(4.88)	(5.10)	(5.64)	(6.89)	(1.51)
Trade Size	Coef.	0.185**	0.586**	0.428**	0.243**	0.068**	0.107**	0.250**	0.338**	0.402**
	t-stat.	(58.72)	(21.19)	(22.04)	(20.83)	(18.91)	(32.43)	(19.63)	(17.20)	(9.12)
Trade Size Squared	Coef.	-0.010**	-0.046**	-0.027**	-0.015**	-0.003**	-0.005**	-0.016**	-0.024**	-0.031**
	t-stat.	(-63.54)	(-22.06)	(-22.17)	(-20.71)	(-17.38)	(-31.91)	(-19.45)	(-18.38)	(-10.08)
SameSide BBO Depth	Coef.	0.0007	0.041**	0.033**	0.029**	0.0004*	0.0005	0.033**	0.028**	0.013**
	t-stat.	(1.54)	(15.97)	(14.98)	(22.22)	(1.97)	(1.71)	(22.19)	(11.49)	(3.00)
OtherSide BBO Depth	Coef.	0.000	-0.068**	-0.0006	-0.031**	0.00001	0.000	-0.0008	-0.012**	-0.033**
	t-stat.	(1.08)	(-22.20)	(-0.90)	(-11.55)	(1.01)	(1.00)	(-0.94)	(-5.51)	(-8.92)
FirstTrade	Coef.	0.171**	0.302**	0.183**	0.089**	0.131**	0.123**	0.121**	0.277**	0.316**
	t-stat.	(15.54)	(9.61)	(7.52)	(4.56)	(8.07)	(8.52)	(6.34)	(10.55)	(7.06)
Duration	Coef.	-0.011**	0.064**	-0.0025	-0.021**	-0.012**	-0.015**	-0.018	0.043**	0.088**
	t-stat.	(-9.64)	(7.78)	(-0.48)	(-5.06)	(-11.78)	(-13.96)	(-3.54)	(6.21)	(6.40)
Duration Squared	Coef.	0.001**	-0.007**	-0.0003	0.002**	0.001**	0.001**	0.001*	-0.004**	-0.009**
	t-stat.	(4.12)	(-8.87)	(-0.50)	(3.66)	(4.80)	(8.33)	2.31	(-6.74)	(-7.23)
Spread	Coef.	13.56**	12.92**	15.40**	10.41**	15.45**	14.37**	11.45**	11.60**	14.78**
	t-stat.	(23.55)	(17.00)	(28.11)	(8.90)	(19.61)	(21.48)	(7.92)	(15.47)	(16.30)
Spread Squared	Coef.	-34.87**	-43.03**	-49.22**	21.06	-40.26	-11.48	14.87	-25.60*	-51.57**
	t-stat.	(-4.06)	(-5.02)	(-6.22)	(0.90)	(-1.69)	(-0.47)	(0.54)	(-2.42)	(-6.08)
Market Trade	Coef.	0.257**	0.877**	0.554**	0.390**	0.128**	0.150**	0.513**	0.771**	1.277**
	t-stat.	(172.98)	(77.14)	(91.78)	(71.00)	(97.26)	(111.01)	(88.66)	(86.69)	(73.49)
Large Trade	Coef.	0.282**	1.203**	0.747**	0.509**	0.115**	0.147**	0.706**	1.135**	1.822**
	t-stat.	(185.58)	(86.13)	(101.27)	(78.12)	(85.52)	(108.20)	(102.41)	(92.44)	(72.97)
# Obs.		2141485	107369	207780	287781	1538555	1750620	240226	105419	45220
R^2		0.108	0.219	0.140	0.104	0.041	0.044	0.111	0.214	0.261

Table 7
Broker Clienteles by Trade Size

This table shows the percentage of trading in each trade-size quartile that the different investor types send through each broker group. We consolidate all investors into three types: (i) all domestic households, (ii) foreign investors who trade via nominee accounts, and (iii) domestic institutional investors (i.e., finance and insurance institutions, non-financial corporations, nonprofits, and government). We use a simple rule to categorize brokerage firms: A broker belongs to a Broker Group that is associated with a certain investor type (households, foreigners, or domestic institutions) if more than 50 percent of the trades that the broker executes involve this investor type. This rule is applied on a stock-by-stock basis because some brokers could see the majority of their orders coming from one investor type (e.g., domestic institutions) in large and active stocks, while having most of their trading in thinly traded stocks come from another investor type (e.g., households). We then put all trades into four trade-size quartiles based on stock-specific breakpoints (i.e., each stock contributes observations to all trade-size quartiles). For each broker group, the cells of the table report the fraction of trades in each trade-size quartile that comes from the actual investor types. We sum the total number of trades in each category for all stocks, and compute the fractions from these aggregate numbers.

Broker Group	**Investor Types**	**Trade Size Quartiles**			
		Q1(smallest)	Q2	Q3	Q4 (largest)
Broker Group 1	Households	81.7%	70.7%	56.6%	40.0%
	Foreigners	4.6%	6.9%	10.5%	16.8%
	Domestic Institutions	13.7%	22.3%	32.9%	43.3%
Broker Group 2	Households	7.9%	1.5%	0.8%	0.4%
	Foreigners	75.9%	81.9%	82.4%	83.6%
	Domestic Institutions	16.2%	16.6%	16.9%	16.0%
Broker Group 3	Households	19.6%	7.2%	5.5%	3.2%
	Foreigners	19.5%	13.5%	15.0%	17.9%
	Domestic Institutions	60.9%	79.2%	79.6%	78.9%

Table 8
Price Impact Regressions with Interactions

This table presents the permanent price impact regressions with interactions between various explanatory variables and the broker group dummy variables (in Panel A) or customer base fractions (in Panel B). The permanent price impact (ppi) is defined as the signed log-change in the midpoint of the quote from an instant before the transaction to five minutes after the transaction. The price impact measure is standardized by subtracting the mean price impact in the same stock for marketable orders in the same direction (i.e., separate means are computed for buys and sells). For ease of interpretation, the permanent price impact measure is multiplied by 100, so a value of "1" indicates that the (standardized) log-change in the quote midpoint is 1%. We estimate the following pooled regressions with stock fixed effects where the dependent variable is the permanent price impact of trades:

$$\begin{aligned}
\text{ppi}_{i,t} = {} & b_1\text{Volume}_{i,t} + b_2\text{SgnReturn}_{i,t} + b_3\text{Volatility}_{i,t} + b_4\text{TradeSize}_{i,t} + b_5\text{TradeSize}^2_{i,t} + b_6\text{SameSideDepth}_{i,t} + b_7\text{OtherSideDepth}_{i,t} + \\
& b_8\text{FirstTrade}_{i,t} + b_9\text{Duration}_{i,t} + b_{10}\text{Duration}^2_{i,t} + b_{11}\text{Spread}_{i,t} + b_{12}\text{Spread}^2_{i,t} + b_{13}\text{MarketTrade}_{i,t} + b_{14}\text{LargeTrade}_{i,t} + \\
& \text{BG}_{i,t} \times \begin{pmatrix} c_0 + c_1\text{Volume}_{i,t} + c_2\text{SgnReturn}_{i,t} + c_3\text{Volatility}_{i,t} + c_4\text{TradeSize}_{i,t} + c_5\text{TradeSize}^2_{i,t} + \\ c_6\text{SameSideDepth}_{i,t} + c_7\text{OtherSideDepth}_{i,t} + c_8\text{FirstTrade}_{i,t} + c_9\text{Duration}_{i,t} + c_{10}\text{Duration}^2_{i,t} + \\ c_{11}\text{Spread}_{i,t} + c_{12}\text{Spread}^2_{i,t} + c_{13}\text{MarketTrade}_{i,t} + c_{14}\text{LargeTrade}_{i,t} \end{pmatrix} + a_i\text{StockDummies}_{i,t} + \varepsilon_{i,t}
\end{aligned}$$

We run the regression separately for one broker group (and its interactions) at a time to simplify the exposition. The indices (i,t) refer to trade t in stock i. In Panel A, *BG* is the dummy variable for the broker group. For ease of presentation we provide the coefficients from each regression in two columns (with the same heading): The first column has the coefficients on the order and market environment attributes, while the second column has the coefficients on the interactions between these attributes and the broker group dummy variable. In Panel B, we replace the broker group dummy variable with a continuous variable (Inv) that represents the fraction of a certain investor type in the customer base of a broker. In the regression labeled "Households," for example, we compute for each broker in each stock the fraction of trades executed by households over the sample period. Hence, the number of distinct values this variable takes in the pooled regression is the number of brokerage firms times the number of stocks. The higher the value of this variable is the greater is the likelihood that an order comes from a household rather than a domestic institution or a foreigner. We follow a similar construction of the variable in the foreigners and domestic institutions regressions. The other variables in the regression belong to three categories. The first category describes recent market activity and includes volume, signed return, and volatility in the five minutes prior to the trade, as well as log duration and squared log duration. FirstTrade is a dummy variable that is set to one for the first trade of the day where we are unable to compute duration. The Duration variable is set to zero in these cases, and therefore the dummy variable FirstTrade shifts the intercept appropriately. The second category describes the state of the limit order book and includes depth at the best bid and offer (same-side and other-side depth relative to the direction of the marketable order), as well as log bid-ask spread and squared log spread. The third category consists of measures of the size of the marketable order: (i) trade size, (ii) trade size squared, (iii) Market Trade dummy (which is equal to one if the size of the marketable order is equal to the number of shares available at the best price in the book), and (iv) Large Trade dummy (which is equal to one if the size of the marketable order is larger than the number of shares available at the best price in the book). We use "**" to indicate significance at the 1% level and "*" to indicate significance at the 5% level (both against a two-sided alternative).

Panel A: Regressions with the Broker Group Dummy Variable

		BG1	**BG2**	**BG3**	Interactions:		**BG1**	**BG2**	**BG3**
					BG	Coef.	-0.947**	0.823**	0.535**
						t-stat.	(-24.54)	(24.99)	(6.83)
Volume	Coef.	0.118**	0.171**	0.168**	BG*	Coef.	0.065**	-0.062**	-0.045**
	t-stat.	(30.19)	(40.40)	(48.59)	Volume	t-stat.	(10.73)	(-10.82)	(-5.38)
SgnReturn	Coef.	-0.065**	-0.057**	-0.060**	BG*	Coef.	0.009	-0.008	-0.003
	t-stat.	(-21.76)	(-16.85)	(-23.55)	SgnReturn	t-stat.	(1.91)	(-1.77)	(-0.35)
Volatility	Coef.	0.031	0.019**	0.021**	BG*	Coef.	-0.018**	0.007	0.024**
	t-stat.	(8.98)	(5.35)	(7.61)	Volatility	t-stat.	(-3.58)	(1.45)	(2.81)
Trade Size	Coef.	0.049**	0.212**	0.194**	BG*	Coef.	0.207**	-0.169**	-0.102**
	t-stat.	(10.21)	(45.66)	(62.19)	Trade Size	t-stat.	(26.14)	(-26.39)	(-6.85)
Trade Size	Coef.	-0.003**	-0.012**	-0.010**	BG*squared	Coef.	-0.012**	0.009**	0.005
Squared	t-stat.	(-11.95)	(-47.97)	(-66.06)	Trade Size	t-stat.	(-29.02)	(29.18)	(6.85)
SameSide	Coef.	0.012**	0.0005	0.0006	BG*Depth	Coef.	-0.012**	0.012**	0.013**
BBO Depth	t-stat.	(10.58)	(1.71)	(1.58)	SameSide	t-stat.	(-9.89)	(8.09)	(7.77)
OtherSide	Coef.	0.00001	0.00004	0.0000	BG*Depth	Coef.	-0.00006	-0.00003	0.002*
BBO Depth	t-stat.	(1.03)	(0.50)	(0.91)	OtherSide	t-stat.	(-0.88)	(-0.45)	(2.13)
FirstTrade	Coef.	0.249**	0.157**	0.160**	BG*	Coef.	-0.100**	0.066**	0.084*
	t-stat.	(13.64)	(12.32)	(13.81)	FirstTrade	t-stat.	(-4.40)	(2.66)	(2.46)
Duration	Coef.	-0.015**	-0.004*	-0.009**	BG*	Coef.	0.023**	-0.010**	-0.012**
	t-stat.	(-12.47)	(-2.01)	(-7.82)	Duration	t-stat.	(9.12)	(-4.20)	(-3.82)
Duration	Coef.	0.001**	-0.0004	0.0005**	BG*squared	Coef.	-0.003**	0.002**	0.001*
Squared	t-stat.	(7.54)	(-1.90)	(3.09)	Duration	t-stat.	(-10.32)	(5.88)	(2.28)
Spread	Coef.	22.036**	11.49**	12.544**	BG*	Coef.	-12.62**	10.16**	10.354**
	t-stat.	(17.77)	(32.71)	(20.84)	Spread	t-stat.	(-10.13)	(7.76)	(8.29)
Spread	Coef.	-38.505	-29.24**	-31.41**	BG*squared	Coef.	18.742	7.098	-40.105
Squared	t-stat.	(-1.75)	(-5.61)	(-3.48)	Spread	t-stat.	(0.84)	(0.27)	(-1.95)
Market	Coef.	0.165**	0.391**	0.267**	BG*	Coef.	0.286**	-0.241**	-0.046**
Trade	t-stat.	(108.57)	(144.82)	(167.42)	MarketTrade	t-stat.	(73.27)	(-75.51)	(-10.42)
Large	Coef.	0.166**	0.454**	0.289**	BG*	Coef.	0.376**	-0.302**	-0.053**
Trade	t-stat.	(111.56)	(154.40)	(177.55)	Large Trade	t-stat.	(92.12)	(-90.73)	(-10.93)
					# Obs.		2141485	2141485	2141485
					R^2		0.120	0.116	0.109

Panel B: Regressions with the Customer Base Continuous Variable

		Households	**Foreigners**	**Domestic Institutions**	Interactions:		**Households**	**Foreigners**	**Domestic Institutions**
					Inv	Coef.	-1.188**	0.976**	0.752**
						t-stat.	(-22.51)	(23.24)	(6.91)
Volume	Coef.	0.105**	0.176**	0.190**	Inv*	Coef.	0.101**	-0.089**	-0.101**
	t-stat.	(25.09)	(38.82)	(35.05)	Volume	t-stat.	(11.51)	(-11.29)	(-6.67)
SgnReturn	Coef.	-0.061**	-0.058**	-0.064**	Inv*	Coef.	0.00297	-0.003	0.011
	t-stat.	(-19.46)	(-16.09)	(-17.51)	SgnReturn	t-stat.	(0.44)	(-0.56)	(1.01)
Volatility	Coef.	0.028**	0.018**	0.020**	Inv*	Coef.	-0.016*	0.009	0.016
	t-stat.	(7.69)	(4.87)	(5.00)	Volatility	t-stat.	(-2.30)	(1.37)	(1.22)
Trade Size	Coef.	0.045**	0.218**	0.221**	Inv*	Coef.	0.258**	-0.203**	-0.140**
	t-stat.	(9.27)	(45.49)	(41.65)	Trade Size	t-stat.	(24.06)	(-25.37)	(-7.00)
Trade Size	Coef.	-0.002**	-0.012**	-0.011**	Inv*squared	Coef.	-0.016**	0.012**	0.007**
Squared	t-stat.	(-9.93)	(-48.70)	(-45.22)	Trade Size	t-stat.	(-27.58)	(29.21)	(7.13)
SameSide	Coef.	0.013**	0.00021	-0.002**	Inv*Depth	Coef.	-0.017**	0.015**	0.014**
BBO Depth	t-stat.	(10.14)	(0.75)	(-4.21)	SameSide	t-stat.	(-9.53)	(7.78)	(6.72)
OtherSide	Coef.	0.00001	0.00002	0.00000	Inv*Depth	Coef.	-0.00007	-0.00002	0.00003
BBO Depth	t-stat.	(1.17)	(0.67)	(-0.35)	OtherSide	t-stat.	(-0.84)	(-0.59)	(0.80)
FirstTrade	Coef.	0.264**	0.157**	0.105**	Inv*	Coef.	-0.157**	0.069*	0.236**
	t-stat.	(13.41)	(11.10)	(5.39)	FirstTrade	t-stat.	(-4.61)	(2.06)	(3.84)
Duration	Coef.	-0.020**	-0.00086	-0.005**	Inv*	Coef.	0.042**	-0.017**	-0.024**
	t-stat.	(-15.87)	(-0.41)	(-3.37)	Duration	t-stat.	(11.32)	(-5.88)	(-5.15)
Duration	Coef.	0.002**	-0.001**	0.00015	Inv*squared	Coef.	-0.006**	0.003**	0.002**
Squared	t-stat.	(10.98)	(-3.51)	(0.67)	Duration	t-stat.	(-12.83)	(7.68)	(2.66)
Spread	Coef.	23.47**	10.626**	9.121**	Inv*	Coef.	-18.798**	13.923**	17.872**
	t-stat.	(18.14)	(27.03)	(8.93)	Spread	t-stat.	(-11.21)	(8.07)	(6.22)
Spread	Coef.	-33.935	-26.413**	-20.495	Inv*squared	Coef.	25.179	8.728	-55.338
Squared	t-stat.	(-1.42)	(-4.61)	(-1.35)	Spread	t-stat.	(0.83)	(0.25)	(-1.25)
Market	Coef.	0.135**	0.418**	0.274**	Inv*	Coef.	0.457**	-0.338**	-0.051
Trade	t-stat.	(83.20)	(139.52)	(121.81)	MarketTrade	t-stat.	(77.80)	(-76.68)	(-7.93)
Large	Coef.	0.132**	0.485**	0.293**	Inv*	Coef.	0.581**	-0.410**	-0.044**
Trade	t-stat.	(84.18)	(151.86)	(135.73)	LargeTrade	t-stat.	(97.34)	(-93.69)	(-7.02)
					# Obs.		2141485	2141485	2141485
					R^2		0.121	0.117	0.109

Figure 1
Success Rate of Database Matching Procedure

This figure looks at the daily success rate of the matching procedure we use to combine our two primary data sources: The Helsinki Stock Exchange (HEX) data and the Finnish Central Securities Depository (FCSD) registry data. The FCSD registry contains the complete trading records of all Finnish investors in domestic publicly-traded stocks. The HEX supervisory files contain detailed information on every order that is entered into the system. Our study requires that we determine which investor type is behind each trade, and therefore we need to match the HEX data with the FCSD registry data. To construct this figure we compute the daily matching rate (the number of trades for which we have identified the investor type behind the trade divided by the total number of trades) for each stock and take the equal-weighted average to find the average matching rate for the day. We then plot the daily matching rate over our sample period, which starts on July 10, 2000 and ends on October 23, 2001.

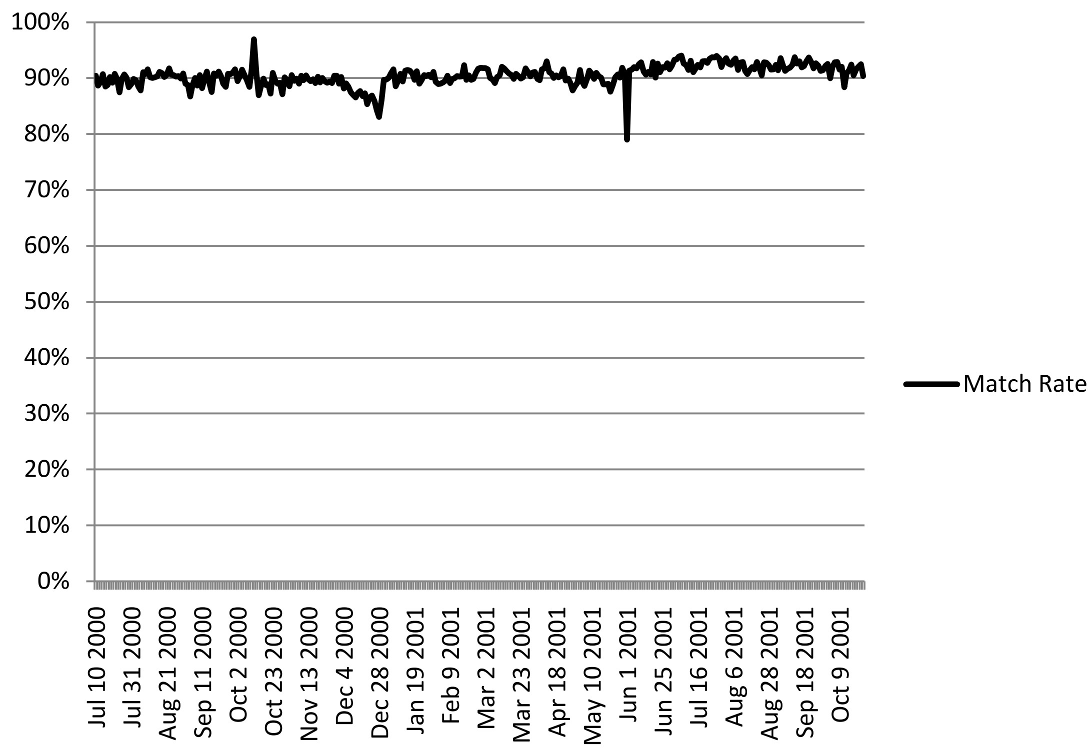

Figure 2
Broker Group Coefficients for Trade-Size Quartiles

This figure presents broker group coefficients from probit regressions that examine whether broker identities provide useful information about investor types over and above the information in order and market environment attributes. We divide the sample of marketable orders into quartiles according to their size, and run separate probit regressions that include stock fixed effects and the same set of explanatory variables as in Table 4 for each trade-size quartile. In the households regressions, the dependent variable is set to one if a household submits the marketable limit order and to zero if either a foreigner or a domestic institution submits the marketable limit order. Similarly, the dependent variable in the foreigners (domestic institutions) regression is set to one if a foreigner (domestic institution) submits the marketable limit order and to zero if the order was submitted by a different type of investor. Explanatory variables consist of broker group dummies (BG1, BG2, and BG3) and a set of control variables that describe recent market activity, the prevailing state of the limit order book, and order attributes. The figure presents the BG1 coefficients from the households regressions in each trade-size quartile, as well as the BG2 coefficients from the foreigners regressions and the BG3 coefficients from the domestic institutions regressions. All of the coefficients that are presented in this figure are statistically significant at the 1% level.

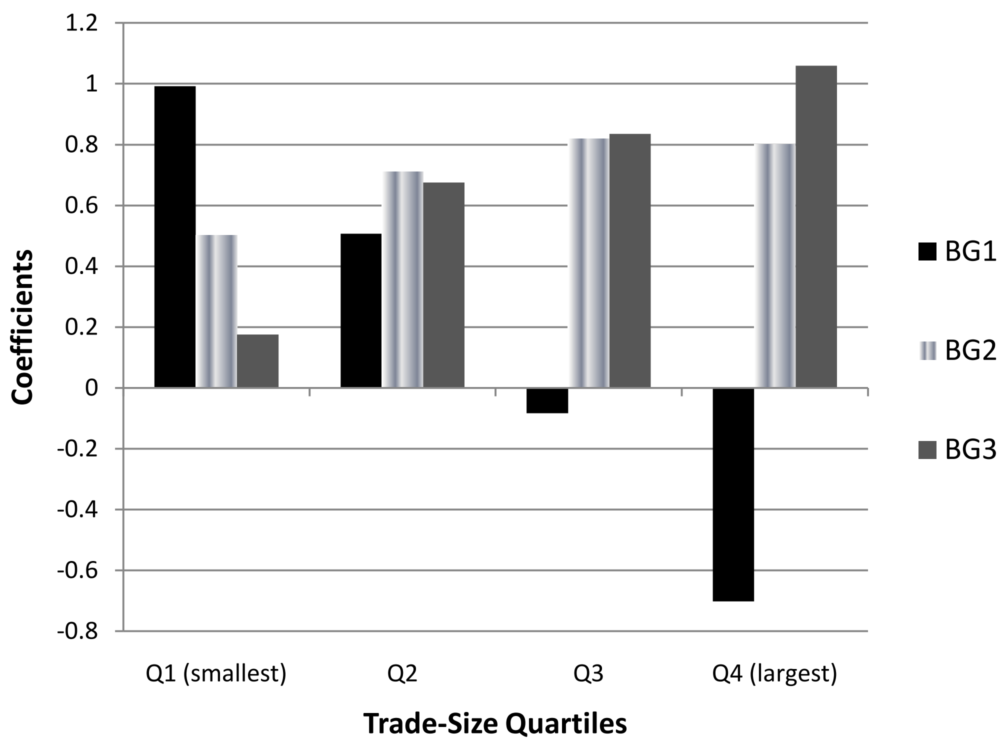

Figure 3
Permanent Price Impact and Trade Size

This figure presents broker group coefficients from the permanent price impact regressions carried out separately on subsamples formed on trade size. We divide the sample of marketable orders into quartiles according to their size, and run separate regressions that include stock fixed effects and the same set of explanatory variables as in Table 6 on each trade-size quartile. The permanent price impact is defined as the signed log-change in the midpoint of the quote from an instant before the transaction to five minutes after the transaction. The price impact measure is standardized by subtracting the mean price impact in the same stock for marketable orders in the same direction (i.e., separate means are computed for buys and sells). For ease of interpretation, the permanent price impact measure is multiplied by 100, so a value of “1” indicates that the (standardized) log-change in the quote midpoint is 1%. Explanatory variables in the regression consist of broker group dummies (BG2 and BG3) and a set of control variables that describe recent market activity, the prevailing state of the limit order book, and order attributes. The figure presents the BG2 and BG3 coefficients from the regression in each trade-size quartile. All of the coefficients that are presented in this figure are statistically significant at the 1% level.