Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market

Alain Chaboud Benjamin Chiquoine Erik Hjalmarsson Clara Vega*

May 13, 2009

Abstract

We study the impact that algorithmic trading, computers directly interfacing with trading platforms, has had on price discovery and volatility in the foreign exchange market, using high frequency data representing a majority of global interdealer trading in three major currency pairs from 2006 to 2007. Our dataset contains precise observations of the size and the direction of the computer-generated and human-generated trades each minute. As such, it allows us to analyze the possible links between algorithmic trading and market volatility and liquidity, to identify whose trades have a more permanent impact on prices, and to study how correlated algorithmic trades are. Our study provides several important insights. First, we observe that algorithmic trades tend to be correlated, suggesting that the algorithmic strategies used in the market are not as diverse as those used by non-algorithmic traders. Second, we find no evident causal relationship between algorithmic trading and increased exchange rate volatility. If anything, the presence of more algorithmic trading is associated with lower volatility. Third, we show that even though some algorithmic traders appear to restrict their activity in the minute following a macroeconomic data release, algorithmic traders increase their provision of liquidity relatively more than non-algorithmic traders over the hour following the release. Fourth, we find that non-algorithmic order flow accounts for most of the (long-run) variance in exchange rate returns, i.e. non-algorithmic traders are better "informed" than algorithmic traders. Fifth, we find evidence that supports the literature that proposes to depart from the prevalent assumption that liquidity providers in limit order books are passive.

*The authors are affiliated with the Division of International Finance at the Federal Reserve Board of Governors. Please address comments to the authors via e-mail at alain.p.chaboud@frb.gov, bchiquoine@gmail.com, erik.hjalmarsson@frb.gov and clara.vega@frb.gov. We are grateful to Terrence Hendershott for his valuable insights, to EBS for providing the data and to Nicholas Klagge and James S. Hebden for their excellent research assistance. We also benefited from the comments of Gordon Bodnar, Luis Marques, Dagfinn Rime, John Schoen, and of participants in the San Francisco AEA 2009 meetings, and the SAIS International Economics Seminar. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

JEL Classification: F3, G12, G14, G15.

Keywords: Algorithmic trading, excess volatility, private information.

1 Introduction

The use of algorithmic trading, where computer algorithms directly manage the trading process at high frequency, has become common in major financial markets in recent years, beginning in the U.S. equity market more than 15 years ago. There has been widespread interest in understanding the potential impact of algorithmic trading on market dynamics, as some analysts have highlighted the potential for improved liquidity and more efficient price discovery while others have expressed concern that it may be a source of increased volatility and reduced liquidity, particularly in times of market stress.[1] Despite this interest, there has been very little formal empirical research on the topic, primarily because of a lack of data where algorithmic trades are clearly identified. A notable exception is a recent paper by Hendershott, Jones, and Menkveld (2007), who get around the data constraint by using the flow of electronic messages on the NYSE as a proxy for algorithmic trading. They conclude that algorithmic trading on the NYSE, contrary to the pessimists' concerns, likely causes an improvement in market liquidity.[2] There has been no formal empirical research on algorithmic trading in the foreign exchange market, where the adoption of algorithmic trading is a far more recent phenomenon than in the equity market, as the two major interdealer electronic trading platforms only began to allow algorithmic trades a few years ago. Growth in algorithmic trading has been rapid, however, and a sizable fraction of foreign exchange transactions currently involve at least one algorithmic counterparty.

In algorithmic trading (AT), computers directly interface with trading platforms, placing orders without human intervention. The computers observe market data and possibly other information at very high frequency, and, based on a built-in algorithm, instantly send back trading instructions. A variety of algorithms are used: some look for arbitrage opportunities, for instance small discrepancies in the exchange rates between three currencies; some seek optimal execution of large orders at the minimum cost; and some seek to implement longer-term trading strategies in search of profits. Among the most recent developments in algorithmic trading, some algorithms now automatically read and interpret economic data releases, generating trading orders before economists have finished reading the first line.[3]

The extreme speed of execution that AT allows and the potential that algorithmic trades may be highly correlated, perhaps as many institutions use similar algorithms, have been cited as reasons for concerns that AT may generate large price swings and market instability. On the other hand, the fact that some algorithms

[1] For instance, an article published by the Financial Times on December 5, 2008, was titled "Algorithmic trades produce snowball effects on volatility."

[2] We also note a paper by Hasbrouck (1996) on program trading, where he analyzes 3 months of data where program trades can be separately identified from other trades. He concludes that both types of orders have an approximately equivalent impact on prices. Algorithmic trading is not exactly equivalent to program trading, though it is a close cousin. In principle, a program trade could be generated by a trader's computer and then the trade conducted manually by a human trader. Our definition of AT refers to the direct interaction of a trader's computer with an electronic trading platform; that is, the automated placement of a trade order on the platform.

[3] The Economist, June 21, 2007

aim for optimal execution at a minimal price impact may be expected to lower volatility. In this paper, we investigate whether algorithmic ("computer") trades and non-algorithmic ("human") trades have different effects on the foreign exchange market. We first ask whether the presence of computer trades causes higher or lower volatility and whether computers increase or reduce liquidity during periods of market stress. We then study the relative importance of human and computer trades in the process of price discovery and re-visit the assumption that liquidity providers are "uninformed."

We formally investigate these issues using a novel dataset consisting of two years (2006-2007) of minute-by-minute trading data from EBS on three exchange rate pairs: the euro-dollar, dollar-yen, and euro-yen. The data represent the vast majority of global interdealer transactions. An important feature of the data is that the volume and direction of human and computer trades each minute are explicitely identified.

We first show some evidence that computer trades are more highly correlated with each other than human trades, suggesting that the strategies used by computers are not as diverse as those used by humans. But the high correlation of computer trades does not seem to necessarily translate into higher volatility. In fact, we find next that there is no evident causal relationship between AT and increased market volatility. If anything, the presence of more algorithmic trading appears to lead to lower market volatility, although the economic magnitude of the effect is small. In our estimations, to account for the potential endogeneity of algorithmic trading with regards to volatility, we instrument for the actual level of algorithmic trading with the installed capacity for algorithmic trading in the EBS system at a given time. We study next the relative provision of market liquidity by computers and humans at the times of the most influential U.S. macroeconomic data release, the nonfarm payroll report. We find that, as a share of total market-making activity, computers tend to pull back slightly at the precise time of the release but then increase their presence in the following hour. This result is robust to considering other important macroeconomic news releases and suggests that computers *do* provide liquidity during periods of market stress.

Finally, we estimate return-order flow dynamics using a structural VAR framework in the tradition of Hasbrouck (1991a). The VAR estimation provides two important insights. First, we find that human order flow accounts for much of the long-run variance in exchange rate returns in the euro-dollar and dollar-yen exchange rate markets, i.e., humans appear to be the "informed" traders in these markets. In contrast, in the euro-yen exchange rate market, computers and humans appear to be equally "informed." In this crossrate, we believe that computers have a clear advantage over humans in detecting and reacting more quickly to triangular arbitrage opportunities, where the euro-yen price is briefly out of line with prices in the euro-dollar and dollar-yen markets. Second, we find that, on average, computers or humans that trade on a price posted by a computer do not impact prices quite as much as they do when they trade on a price posted by a human. One possible interpretation of this result is that computers tend to place limit orders more strategically than humans do. This empirical evidence supports the literature that proposes to depart from the prevalent assumption that liquidity providers in limit order books are passive.[4]

The paper proceeds as follows. In Section 2 we introduce the EBS exchange rate data, describing the evolution over time of algorithmic trading and the pattern of interaction between human and algorithmic traders. In Section 3 we study the correlation of algorithmic trades. In Section 4 we analyze the relationship between algorithmic trading and exchange rate volatility. In Section 5 we discuss the provision of liquidity by computers and humans at the time of a major data release. In Section 6 we report the results of the high-frequency VAR analysis. We conclude in Section 7.

2 Data description

Today, two electronic platforms process the vast majority of global interdealer spot trading in the major currency pairs, one offered by Reuters, and one offered by EBS.[5] These platforms, which are both electronic limit order books, have become essential utilities for the foreign exchange market. Importantly, trading in each major currency pair has over time become very highly concentrated on only one of the two systems. Of the most traded currency pairs, the top two, euro-dollar and dollar-yen, trade primarily on EBS, while the third, sterling-dollar trades primarily on Reuters. As a result, the reference price at any moment for, say, spot euro-dollar, is the current price on the EBS system, and all dealers across the globe base their customer and derivative quotes on that price. EBS controls the network and each of the terminals on which the trading is conducted. Traders can enter trading instructions manually, using an EBS keyboard, or, upon approval by EBS, via a computer directly interfacing with the system. The type of trader (human or computer) behind each trading instruction is recorded by EBS, allowing for our study.[6]

We have access to AT data from EBS from 2003 through 2007. We focus on the sample from 2006 and 2007, because, as we show in Figure 1, algorithmic trades were a very small portion of all trades in the earlier years. In addition to the full 2006-2007 sample, we also consider a sub-sample covering the months of September, October, and November of 2007, when algorithmic trading played an even more important role than earlier in the sample.[7] We study the three most-traded currency pairs on the EBS system: euro-dollar, dollar-yen, and euro-yen.

The quote data, at the one-second frequency, consist of the highest bid quote and the lowest ask quote on

[4] For example, Chakravarty and Holden (1995), Kumar and Seppi (1994), Kaniel and Liu (2006), and Goettler, Parlour and Rajan (2007) allow informed investors to use both limit and market orders. Bloomfield, O'Hara and Saar (2005) argue that informed traders are natural liquidity providers and Angel (1994) and Harris (1998) show that informed investors can optimally use limit orders when private information is sufficiently persistent.

[5] EBS has been part of the ICAP group since 2006.

[6] EBS uses the name "automated interface" (AI) to describe trading activity directly generated by a computer, activity we call AT.

[7] We do not use December 2007 in the sub-sample to avoid the influence of year-end effects.

the EBS system in these currency pairs, from which we construct one-second mid-quote series and compute one-minute exchange rate returns; all the quotes are executable. The transactions data are at the one-minute frequency and provide detailed information on the volume and direction of trades that can be attributed to computers and human in each currency pair. Specifically, the transactions volume data are broken down into categories specifying the "maker" and "taker" of the trade (i.e., human or computer), and the direction of the trade (i.e., buy or sell the base currency), for a total of eight different combinations. That is, the first transaction category may specify, say, the minute-by-minute volume of trade that results from a human taker buying the base currency by "hitting" a quote posted by a human maker. We would record this activity as the human-human buy volume, with the aggressor (taker) of the trade buying the base currency. The human-human sell volume is defined analogously, as are the other six buy and sell volumes that arise from the remaining combinations of computers and humans acting as makers and takers.

From these eight types of buy and sell volumes, we can construct, for each minute, trading volume and order flow measures for each of the four possible pairs of human and computer makers and takers: human-maker/human-taker (HH), computer-maker/human-taker (CH), human-maker/computer-taker (HC), and computer-maker/computer-taker (CC).[8] That is, the *sum* of the buy and sell volumes for each pair gives the volume of trade attributable to that particular combination of maker and taker (which we symbolize as, $Vol(HH)$ or $Vol(HC)$, for example). The *difference* between the buy and sell volume for each pair gives us the order flow attributable to that maker-taker combination (which we symbolize simply as HH or HC, for example). The sum of the four volumes, $Vol(HH + CH + HC + CC)$, gives the total volume of trade in the market. The sum of the four order flows, $HH + CH + HC + CC$, gives the total (market-wide) order flow.[9] Throughout the paper, we will use the expression 'order flow' to refer both to the market-wide order flow and to the order flows from other possible decompositions, with the distinction clearly indicated. Importantly, the data allows us to consider order flow broken down by the type of trader who initiated the trade, human-taker order flow ($HH + CH$) and computer-taker order flow ($HC + CC$).

The main goal of this paper is to analyze the effect algorithmic trading has on price discovery and volatility in the foreign exchange market. As we show later, in our exchange rate data as in other financial data, the net of signed trades from the point of view of the takers (the market-wide order flow) is highly positively correlated with exchange rate returns, so that the takers are considered to be more "informed" than the makers. Thus, in our analysis of the relative effects of human and computer trades in the market, we consider prominently the order flow decomposition into human-taker order flow and computer-taker order flow. However, we also

[8] The naming convention for "maker and taker" reflects the fact that the "maker" posts quotes before the "taker" accepts to trade at that price. Posting quotes is, of course, the traditional role of the market-"maker."

[9] There is a very high correlation in this market between trading volume per unit of time and the number of transactions per unit of time, and the ratio between the two does not vary much over our sample. Order flow measures based on amounts transacted and those based on number of trades are therefore very similar.

consider two other decompositions in our work. We consider the most disaggregated decomposition of order flow (HH, CH, HC, CC), as this decomposition allows us to study whether the liquidity suppliers, who are traditionally assumed to be "uninformed", are posting quotes strategically. This situation is more likely to arise in our database, a pure limit order book market, than in a hybrid market like the NYSE, because, as Parlour and Seppi (2008) point out, the distinction between liquidity supply and liquidity demand in limit order books is blurry.[10] We also decompose the data by maker type (human or computer) in order to study whether computers or humans are providing liquidity during the release of public information, which are periods of high exchange rate volatility or market stress.

In our analysis, we exclude data collected from Friday 17:00 through Sunday 17:00 New York time from our sample, as activity on the system during these "non-standard" hours is minimal and not encouraged by the foreign exchange community. We also drop certain holidays and days of unusually light volume: December 24-December 26, December 31-January 2, Good Friday, Easter Monday, Memorial Day, Labor Day, Thanksgiving and the following day, and July 4 (or, if this is on a weekend, the day on which the U.S. Independence Day holiday is observed).

We show summary statistics for the one-minute returns and order flow data in Table 1. This table contains a number of noteworthy features. First, order flow, whether in total, broken down by human and computer takers, or broken down into the 4 possible pairs of makers and takers, is serially positively correlated, which is consistent with some informed trading models. For example, Easley and O'Hara (1987) model a situation where sequences of large purchases (sales) arise when insiders with positive (negative) signals are present in the market. He and Wang (1995) also show that insiders with good (bad) news tend to buy (sell) repeatedly until their private information is revealed in the prices. The positive serial correlation in order flow is also consistent with strategic order splitting, i.e. a trader willing to buy for informational or non-informational reasons and splitting his order to reduce market impact. Second, the standard deviations of the various order flows differ by exchange rates, by type of taker and across maker/taker pairs. These differences will be important in the interpretation of the upcoming VAR analysis and variance decomposition.

We show in Figure 1, from 2003 through 2007 for our three major currency pairs, the fraction of trading volume where at least one of the two counterparties was an algorithmic trader, $Vol(CH + HC + CC)$ as a fraction of total volume. From its beginning in the second half of 2003, the fraction of trading volume involving AT grew by the end of 2007 to near 60% for euro-dollar, and dollar-yen trading, and to about 80% for euro-yen. Figure 2 shows, for our three currency pairs, the evolution over time of the four different possible types of trades (i.e. $Vol(HH)$, $Vol(CH)$, $Vol(HC)$, and $Vol(CC)$, as fractions of the total volume).

[10] Parlour and Seppi (2008) note that in a limit order book investors with active trading motives, some of which are "informed" traders, may choose to post limit orders that are more aggresive than those a disinterested liquidity provider would use but less aggresive than market orders.

By the end of 2007, in the euro-dollar and dollar-yen markets, human to human trades, in black, accounted for slightly less than half of the volume, and computer to computer trades, in green, for about ten to fifteen percent. These percentages reflect, in part, the fact that there are more human participants in the market than computer participants. In euro-dollar and dollar-yen, we note that $Vol(HC)$ and $Vol(CH)$ are about equal to each other, i.e. computers "take" prices posted by humans, in red, about as often as humans take prices posted by market-making computers, in blue. The story is different for the cross-rate, the euro-yen currency pair. By the end of 2007, there were more computer to computer trades than human to human trades. But the most common type of trade was computers trading on prices posted by humans. We believe this reflects computers taking advantage of short-lived triangular arbitrage opportunities, where prices set in the euro-dollar and dollar-yen markets are very briefly out of line with the euro-yen cross rate. In interpreting our results later in the paper, we will keep in mind that trading volume is largest in the euro-dollar and dollar-yen markets, and price discovery happens mostly in those markets, not in the cross-rate. Our conclusions based on the euro-dollar and dollar-yen markets wil then be more easily generalized than those based on the euro-yen market. Table 2 tabulates the averages of the volume fractions shown in Figures 1 and 2, both for the full 2006-2007 sample and the shorter three-month sub-sample.

3 How Correlated Are Algorithmic Trades and Strategies?

We first investigate the proposition that computers tend to have trading strategies that are more correlated than those of humans. Since the outset of the financial turmoil in the summer of 2007, multiple articles in the financial press have suggested that AT programs tend to be similarly designed, leading them to take the same side of the market in times of high volatility, and potentially exaggerating market movements.

One such instance may have happened on August 16, 2007, a day of extreme volatility in dollar-yen, the highest in that currency pair over our sample period. On that day, the Japanese yen appreciated sharply against the U.S. dollar around 6:00 a.m. and 12:00 p.m. (NY time), as shown in Figure 3. The figure also shows, for each 30-minute interval in the day, computer-taker order flow ($HC + CC$) in the top panel and human-taker order flow ($HH + CH$) in the lower panel. The two sharp exchange rate movements mentioned happened when computers, as a group, aggressively sold dollars and bought yen. We note that computers, during these episodes, mainly trading with humans not with other computers. Human order flow at those times was, in contrast, quite small, even though the overall trading volume initiated by humans (not shown) was well above that initiated by computers (human takers were therefore selling and buying dollars in fairly equal amounts). The "taking" orders generated by computers were then far more correlated than the taking orders generated by humans. After 12:00 p.m., humans traders , as a group, then bought dollars fairly aggressively, and the appreciation of the yen against the dollar was partially reversed. This is only a single example, of course, but it leads us to ask how correlated computer trades and strategies have tended to be overall.

We do not know precisely the exact mix of the various strategies used by algorithmic traders on EBS. Traders keep the information about their own strategies confidential, including, to some extent, from EBS, and EBS also keeps what they know confidential. However, one can get a general sense of the market and of the strategies in conversations with market participants. About half of the algorithmic trading volume on EBS is believed to come from what is often known as the "professional trading community," which primarily refers to hedge funds and commodity trading advisors (CTAs). These participants, until very recently, could not trade manually on EBS, so all their trades were algorithmic. Some hedge funds and CTAs trade at very high frequency as they seek to exploit short-lived arbitrage opportunities, including triangular arbitrage, often accessing several trading platforms. Others implement lower-frequency strategies, often grouped under the 'statistical arbitrage' appellation, including carry trades, momentum trades, and strategies spanning several asset classes. Only a very small fraction of the trading volume in our sample period is believed to have been generated by algorithms designed to quickly react to data releases. The other half of the algorithmic trading volume comes from foreign exchange dealing banks, the only participants allowed on the EBS system until early this decade. Some of the banks' algorithmic trading is clearly related to activity on their own customer-to-dealer platforms, to automated hedging activity, and to minimizing the impact of the execution of large orders. But a sizable fraction is believed to be proprietary trading implemented algorithmically, likely using a mix of strategies similar to those employed by hedge funds and CTAs. Overall, market participants generally believe that the mix of algorithmic strategies used in the foreign exchange market differs from that seen in the equity market, where optimal execution algorithms seem to be more prevalent.

The August 16, 2007 episode shown above, where sharp movements in dollar-yen were clearly associated with algorithmic activity, was widely viewed as the result of a sudden unwinding of the yen-carry trade, with hedge funds and proprietary trading desks at banks rushing to close risky positions and buying yen to pay back low-interest loans. The evidence in this episode raises the possibility that many algorithmic traders were using fairly similar carry trade and momentum strategies at the time, leading to the high correlation of algorithmic orders and to sharp exchange rate movements. But this is only one episode in our two-year sample. Next, we investigate whether there is evidence that, over the entire sample, the strategies used by algorithmic traders have tended to be more correlated than those used by human traders.

If computers and humans are indifferent between taking or making liquidity at a given point in time, then we should observe that computers and humans trade with each other in proportion to their relative presence in the market. If, on the other hand, computers tend to have more homogeneous trading strategies, we should observe computers trading less among themselves and more with humans. At the extreme, if all computers used the very same algorithms and had the exact same speed of execution, we would observe no trading volume among computers. Therefore, the fraction of trades conducted between computers contains information on how correlated their strategies are.

We consider a simple model in which there are H_m potential human-makers (the number of humans that are standing ready to provide liquidity), H_t potential human-takers, C_m potential computer-makers, and C_t potential computer-makers. At a given period of time the probability of a computer providing liquidity to a trader is equal to $Prob(computer - make) = \frac{C_m}{C_m + H_m}$, which we label for simplicity as α_m, and the probability of a computer taking liquidity from the market is $Prob(computer - take) = \frac{C_t}{C_t + H_t} = \alpha_t$. The remaining makers and takers are humans, in proportions $(1 - \alpha_m)$ and $(1 - \alpha_t)$, respectively. Assuming that these events are independent the probabilities of the four possible trades, human-maker/human-taker, computer-maker/human-taker, human-maker/computer-taker and computer-maker/computer taker, are:

$$\begin{aligned} Prob(HH) &= (1 - \alpha_m)(1 - \alpha_t) \\ Prob(HC) &= (1 - \alpha_m)\alpha_t \\ Prob(CH) &= \alpha_m(1 - \alpha_t) \\ Prob(CC) &= \alpha_m \alpha_t. \end{aligned}$$

These probabilities yield the following identity,

$$Prob(HH) \times Prob(CC) \equiv Prob(HC) \times Prob(CH),$$

which can be re-written as,

$$\frac{Prob(HH)}{Prob(CH)} \equiv \frac{Prob(HC)}{Prob(CC)}.$$

We label the first ratio, $RH \equiv \frac{Prob(HH)}{Prob(CH)}$, the "human-taker" ratio and the second ratio, $RC \equiv \frac{Prob(HC)}{Prob(CC)}$, the "computer-taker" ratio. In a world with more human traders (both makers and takers) than computer traders, each of these ratios will be greater than one, because $Prob(HH) > Prob(CH)$ and $Prob(HC) > Prob(CC)$ i.e., computers take liquidity more from humans than from other computers, and humans take liquidity more from humans than from computers. However, under the baseline assumptions of our random-matching model, the identity shown above states that the ratio of ratios, $R \equiv \frac{RC_k}{RH_k}$, will be equal to one. In other words, humans will take liquidity from other humans in a similar proportion that computers take liquidity from humans.

Turning to the data, under the assumption that potential human-takers are randomly matched with potential human-makers i.e., that the probability of a human-maker/human-taker trade is equal to the one predicted by our model, $Prob(HH) = \frac{H_m \times H_t}{(H_m + C_m) \times (H_t + C_t)}$, we can now derive implications from observations of R, our ratio of ratios. In particular, finding that $R > 1$ must imply that algorithmic strategies are more correlated than what our random matching model implies. In other words, for $R > 1$ we must observe that either computers trade with each other less than expected ($Prob(CC) < \frac{C_m \times C_t}{(H_m + C_m) \times (H_t + C_t)}$) or that computers trade with humans more than expected (either $Prob(CH) > \frac{C_m \times H_t}{(H_m + C_m) \times (H_t + C_t)}$ or $Prob(HC) > \frac{H_m \times C_t}{(H_m + C_m) \times (H_t + C_t)}$).

Our dataset allows us to estimate an ex-post proxy for R. Namely, for each trading day we estimate $\widehat{RH} = \frac{Vol(HH)}{Vol(CH)}$ and $\widehat{RC} = \frac{Vol(HC)}{Vol(CC)}$, where $Vol\,(HH)$ is the daily trading volume between human makers and human takers, and so forth. In Table 3 we show the mean of the daily ratio of ratios, $\widehat{R} = \frac{\widehat{RC}}{\widehat{RH}}$, for each currency pair for the full sample and the three-month sub-sample. In contrast to the above theoretical prediction that $R \equiv \frac{RC}{RH} = 1$, we find that for all currency pairs $\widehat{R}$ is statistically greater than one. This result is very robust: in euro-dollar, all daily observations of $\widehat{R}$ are above one, and only a very small fraction of the daily observations are below one for the other currency pairs. The results then show that computers do not trade with each other as much as random matching would predict. We take this as evidence that algorithmic strategies are likely less diverse than the trading strategies used by human traders.

This finding, combined with the observed growth in algorithmic trading over time, may raise some concerns about the impact of AT on volatility in the foreign exchange market. As mentioned earlier, some articles in the financial press have pointed to the possible danger of having many algorithmic traders take the same side of the market at the same moment. However, we note that, in principle, a high correlation of algorithmic strategies does not necessarily lead to higher volatility or large swings in exchange rates. If, as during the episode shown at the beginning of the section, the high correlation reflects a high number of traders using the same carry trade or momentum strategies, then there may be reason for concern. However, if, for instance, many algorithmic traders use similar triangular arbitrage strategies, trading mainly on quotes posted by humans and little with other computers, the high correlation of those strategies should have no impact on volatility, and may even lower volatility as it improves the efficiency of the price discovery process. Next, we explicitly investigate the relationship between the presence of algorithmic trading and market volatility.

4 The impact of algorithmic trading on volatility

In this section, we attempt to estimate whether the presence of algorithmic trading causes disruptive market behavior in the form of increased volatility. In particular, we test for a causal relationship between the fraction of daily algorithmic trading, relative to the overall daily volume, and daily realized volatility.

4.1 Identification

The main challenge in identifying a causal relationship between algorithmic trading and volatility is the potential endogeneity of algorithmic trading. That is, although one may conjecture that algorithmic trading impacts volatility, it is also plausible that algorithmic trading activity may be a function of the level of volatility. For instance, highly volatile markets may present comparative advantages to automated trading algorithms relative to human traders, which might increase the fraction of algorithmic trading during volatile periods. In contrast, however, one could also argue that a high level of volatility might reduce the informativeness of historical price patterns on which some trading algorithms are likely to base their decisions, and thus reduce the effectiveness of the algorithms and lead them to trade less. Thus, one can not easily determine in what direction the bias will go in an OLS regression of volatility on the fraction of algorithmic trading. To deal with the endogeneity issue, we adopt an instrumental variable (IV) approach as outlined below.

We are interested in estimating the following regression equation,

$$RV_{i,t} = \alpha_i + \beta_i AT_{i,t} + \gamma_i' \tau_{i,t} + \sum_{k=1}^{22} \delta_i RV_{i,t-k} + \epsilon_{i,t}, \tag{1}$$

where $i = 1, ..., 5$ represents currency pairs and $t = 1, ..., T$, represents time. $RV_{i,t}$ is (log) realized daily volatility, $AT_{i,t}$ is algorithmic trading volume at time t in currency pair i , $\tau_{i,t}$ is either a time trend or a set of time dummies that control for secular trends in the data, and $\epsilon_{i,t}$ is an error term that is assumed to be uncorrelated with $RV_{i,t-k}$, $k \geq 1$, but not necessarily with $AT_{i,t}$. The large number of lags of volatility, which covers the business days of the past month, is included to control for the strong serial correlation in volatility (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003 and Bollerslev and Wright, 2000). The exact definitions of $RV_{i,t}$, $AT_{i,t}$, and $\tau_{i,t}$ are given below.

The main focus of interest is the parameter β_i, which measures the impact of algorithmic trading on volatility in currency pair i. However, since $AT_{i,t}$ and $\epsilon_{i,t}$ may be correlated, due to the potential endogeneity discussed above, the OLS estimator of β_i may be biased. In order to obtain an unbiased estimate, we will therefore consider an instrumental variable approach. Formally, we need to find a variable, or set of variables, $z_{i,t}$, that is uncorrelated with $\epsilon_{i,t}$ (validity of the instrument) and correlated with $AT_{i,t}$ (relevance of the instrument).

The instrument we propose to use is the fraction of algorithmic trading terminals relative to the total number of trading terminals in the EBS system.[11] That is, in order to place algorithmic trades, a special user interface is required, and the total number of such user interfaces thus provides a measure of the overall algorithmic trading 'capacity' in the market. The ratio of these algorithmic trading terminals to the total number of trading terminals therefore provides a measure of the potential fraction of algorithmic trading. Since the number of trading terminals of each type should clearly be exogenous with regards to daily market volatility, the fraction of AT terminals provides a valid instrument. In addition, it is positively correlated with the fraction of algorithmic trading and generally provides a relevant instrument as seen from the tests for weak instruments discussed below.

Under the breakdown provided by our data, there are three types of terminals in the EBS system: purely algorithmic trading terminals, purely manual trading terminals, and dual trading terminals that can handle both manual and algorithmic trades. We consider two natural instrumental variables: the fraction of pure AT terminals over the total number of terminals (including pure AT, manual, and dual ones), and the fraction of the sum of pure AT *and* dual terminals over the total number of terminals. Since it is not obvious which variable is the better instrument, we use both simultaneously.[12]

The data on AT terminals is provided on a monthly basis, whereas the data on realized volatility and algorithmic trading are sampled on a daily frequency. We therefore transform the AT terminals data to daily data by repeating the monthly value each day of the month. Although this leads to a dataset of two years of daily data, the number of daily observations (498) will overstate the effective number of observations, since the coefficient on AT participation will be identified from monthly variations in the instrumental variables. Transforming the instruments to a daily frequency is, however, more efficient than transforming all data to a monthly frequency, since the daily data helps to identify the monthly shifts.

The instrumental variable regressions are estimated using Limited Information Maximum Likelihood (LIML), and we test for weak instruments by comparing the first stage $F-$statistic for the excluded instruments to the critical values of Stock and Yogo's (2005) test of weak instruments. We use LIML rather than two-stage least squares since Stock and Yogo (2005) show that the former is much less sensitive to weak instruments than the latter (see also Stock et al., 2002).

[11] More precisely, we actually observe the number of EBS 'deal codes' of each type. Each deal code may correspond to a small trading floor (many with only one terminal) or to part of a larger trading floor, with several terminals. The actual number of terminals and the number of deal codes is highly correlated and for simplicity, we refer to the variable as number of terminals. These data are confidential.

[12] Regressions not reported here show that using the pure AT terminals as a single instrument gives qualitatively similar results to those presented below based on both instruments. Using the fraction of the sum of both pure and dual AT terminals as a single instrument also leads to the same qualititative conclusion, but with more signs of weak instruments.

4.2 Variable definitions

4.2.1 Realized Volatility

Volatility is measured as the *daily realized volatility* obtained from one minute returns; that is, the volatility measure is equal to the daily sum of squared one minute log-price changes. The use of realized volatility, based on high-frequency intra-daily returns, as an estimate of ex-post volatility is now well established and generally considered the most precise and robust way of measuring volatility. Although many older studies relied on five minute returns in order to avoid contamination by market microstructure noise (e.g. Andersen et al., 2001), recent work shows that sampling at the one-minute frequency, or even higher frequencies, does not lead to biases in liquid markets (see, for instance, the results for liquid stocks in Bandi and Russel, 2006, and the study by Chaboud et al., 2007, who explicitly examine EBS data on the euro-dollar exchange rate during 2005 and finds that sampling frequencies upwards of once every 20 seconds does not lead to noticeable biases). Here, we restrict ourselves to using minute-by-minute data.[13] Following the common conventions in the literature on volatility modelling (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003), the realized volatility is log-transformed to obtain a more well behaved time-series.

4.2.2 Algorithmic trading

We consider two measures of the amount of algorithmic trading, $AT_{i,t}$, in a given currency pair: computer-participation volume and computer-taker volume. The first is simply the percent of the overall trading volume that includes an algorithmic trader as either a maker or a taker ($Vol(CH + HC + CC)$); that is, the percent of trading volume where a computer was involved in at least one side of the trade. In addition, we also consider an alternative measure defined as the fraction of overall trading volume that is due to a computer-taker ($Vol(HC + CC)$).

4.2.3 Time controls

As seen in Figure 4, there is a clear secular trend in both the computer-participation and computer-taker volume, which is not present in realized volatility. Euro-dollar, dollar-yen, and euro-yen volatility is trending down at the beginning of the period and starts to trend up in the summer of 2007. In order to control for the trend in algorithmic trading in the regression, we include either a 'linear' quarterly time trend or a full set of year-quarter dummies, one for each year-quarter pair in the data (8 dummies). That is, the linear quarterly time trend stays constant within each quarter and increases by the same amount each quarter, whereas the year-quarter dummies allows for a more flexible trend specification that can shift in arbitrary fashion from

[13] Using realized volatility based on five-minute returns leads to results that are very similar to those reported below for the one-minute returns, and the qualitative conclusions are identical.

year-quarter to year-quarter. Both secular trend specifications are thus fixed within each quarter. This restriction is imposed in order to preserve the identification coming from the monthly instrumental variables. Using monthly, or finer, time dummies would eliminate the variation in the instrument and render the model unidentified. Although it is theoretically possible to include a monthly time trend, this would lead to very weak identification empirically. The monthly frequency of the instrumental variable is clearly a restriction and ideally a more flexible trend specification would be desirable. The quarterly trend or dummies do, however, likely control for most of the secular shifts in algorithmic trading.

4.3 Empirical results

The regression results are presented in Table 4. We present OLS and LIML-IV results, with either the quarterly trend or the year-quarter dummies included. We show in Panels A and B the results for the computer-participation volume, and in Panels C and D the results for computer-taker volume. We report results for the sample starting in January 2006 and ending in December 2007. In order to save space, we only show the estimates of the coefficients in front of the fraction of algorithmic trading volume variables.

The OLS results, which are likely to be biased due to the aforementioned endogeneity issues, show a fairly clear pattern of a positive correlation between volatility and AT participation, with several positive and statistically significant coefficients. The R^2s are fairly large, reflecting the strong serial correlation in realized volatility, which is picked up by the lagged regressors. There are also no systematic differences between the quarterly trend and quarterly dummies specifications.

Turning to the more interesting IV results, which controls for the endogeneity bias, the coefficient estimates change fairly dramatically. All point estimates are now negative and some of them are statistically significant. Thus, if there is a *causal* relationship between the fraction of algorithmic trading and the level of volatility, all evidence suggests that it is negative, such that increased AT lowers the volatility in the market. The stark difference between the IV and OLS results shows the importance of controlling for endogeneity when estimating the causal effect of AT on volatility; the opposite conclusion would have been reached if one ignored the endogeneity issue. The evidence of a statistically significant relationship is fairly weak, however, with most coefficients statistically indistinguishable from zero. The more restrictive quarterly trend specification suggests a significant relationship for the euro-dollar and dollar-yen, but this no longer holds if one allows for year-quarter dummies.

To the extent that the estimated coefficients are statistically significant, it is important to discuss the *economic* magnitude of the estimated relationship between AT and volatility. The regression is run with log volatility rather than actual volatility, which makes it a little less straightforward to interpret the size of the coefficients. However, some back-of-the-envelope calculations can provide a rough idea. Suppose that the coefficient on computer-participation volume is about -0.01, which is in line with the coefficient estimates for the euro-dollar. The average monthly shift in computer-taker volume in the euro-dollar is about 1.5 percentage points and the average log-volatility in the euro-dollar is about 3.76 (with returns calculated in basis points), which implies an annualized volatility of about 6.82 percent. Increasing computer-taker volume by 1.5 percentage points decreases log-volatility by 0.015 and results in an annualized volatility of about 6.72. Thus, a typical change in computer-taker volume might change volatility by about a tenth of a percentage point in annualized terms.

The first stage $F-$statistics for the excluded instruments in the IV regressions are also reported in Panels B and D. Stock and Yogo (2005) show that this $F-$statistic can be used to test for weak instruments. Rejection of the null of weak instruments indicates that standard inference on the IV-estimated coefficients can be performed, whereas a failure to reject indicates possible size distortions in the tests of the LIML coefficients. The critical values of Stock and Yogo (2005) are designed such that they indicate a maximal actual size for a nominal sized five percent test on the coefficient. Thus, in the case considered here with two excluded instruments and one endogenous regressor, a value greater than 8.68 for this $F-$statistic indicates that the maximal size of a nominal 5 percent test will be no greater than 10 percent, which might be deemed acceptable; a value greater than 5.33 for the $F-$statistic indicates a maximal size of 15 percent for a nominal 5 percent test. In general, the larger the $F-$statistic, the stronger the instruments. As is evident from the table, there are no signs of weak instruments in the specification with a quarterly trend. There are, however, signs of weak instruments in the case with year-quarter dummies, for the euro-yen. This is not too surprising given that the instruments only change on a monthly frequency and the year-quarter dummies therefore put a great deal of strain on the identification mechanism. Importantly, though, the results for the two major currency pairs is robust to any weak instrument problems and the reported coefficients and standard errors are unbiased.

To sum up, the evidence of any causal effect of algorithmic trading on volatility is weak, but what evidence there is points fairly consistently towards a negative relationship. There is thus no systematic statistical evidence to back the often voiced opinion that AT leads to increased levels of volatility. If anything, the contrary appears to be true.

5 Who provides liquidity during the release of public announcements?

In the previous section we discuss one of the major concerns regarding algorithmic trading, namely, whether AT causes exchange rate volatility. We now examine another major concern, whether AT improves or reduces liquidity during stress periods when it is needed the most. To answer this question, we cannot simply regress computer-maker volume, a proxy for liquidity provided by computers, on exchange rate volatility, a proxy for stress periods, because, as we discussed in the previous section, volume and volatility are endogenous variables. In contrast to the previous section we do not estimate an IV regression because there are no obvious instruments for volatility.[14] Instead, we follow the event study literature and compare the liquidity provision by humans and computers during U.S. nonfarm payroll announcements, a period of exogenous heightened volatility, to the liquidity provision by both agents during non-announcement days.[15] This comparison will help us determine who provides relatively more liquidity during stress periods.

We consider two liquidity provision estimates: a one-minute estimate and a one-hour estimate. The one-minute estimate is calculated using 8:30 a.m. to 8.31 a.m. ET (when U.S. nonfarm payroll is released) volume observations, while the one-hour estimate is calculated using observations from 8:25 am to 9:24 am ET. We define the one-minute (one-hour) liquidity provision by humans, LH, as the sum of human-maker volume, $Vol(HH+HC)$, divided by total volume during that period, and the one-minute (one-hour) liquidity provision by computers, LC, as the sum of computer-maker volume, $Vol(CC+CH)$, divided by total volume during that period. Similar to the liquidity provision measures, we define the one-minute volatility as the squared 1-minute return from 8:30 a.m. to 8.31 a.m. ET and the one-hour volatility as the sum of squared 1-minute returns from 8:25 am to 9:24 am ET.

As we discussed in the previous section, exchange rate trading volume follows a secular trend. Thus, we cannot compare average liquidity provision during announcement days estimated using our full sample period to average liquidity provision during non-announcement days. Instead, each announcement day t_a we divide the liquidity provision during that day by the liquidity provision surrounding the announcement day. We then average this ratio over our sample period to test the hypothesis that this ratio is equal to one. Specifically, we divide the one-minute (one-hour) liquidity provision by humans, LH_a, (computers,

[14] One could consider macroeconomic news announcements as potential instruments for volatility. However, macroeconomic news announcements are exogeneous variables that cause both foreign exchange rate volatility and liquidity changes. Since we cannot assume that the effect macroeconomic news announcements have on liquidity is only due to the effect macroeconomic news announcements have on volatility, the exclusion restriction required by IV estimation is violated.

[15] Andersen and Bollerslev (1998), among others, refer to the nonfarm payroll report as the "king" of announcements, because of the significant sensitivity of most asset markets to its release. We note that our results are qualitatively similar when we consider other important macroeconomic news announcements. However, the nonfarm payroll announcement is associated with the highest volatility among all the macroeconomic announcements we considered.

LC_a) estimated during announcement day t_a by the one-minute (one-hour) liquidity provision by humans, LH_n, (computers, LC_n) estimated during the surrounding non-announcement day period, t_n, defined as 10 business days before and after the nonfarm payroll release date t_a; the liquidity provision measures on the non-announcement days are calculated in the same manner as on the announcement days, using data only for the periods 8:30 a.m. to 8.31 a.m. ET or 8:25 am to 9:24 am ET for the one-minute and one-hour measures, respectively.[16] This methodology amounts to using a non-parametric approach to de-trend the data. We follow the same procedure with our one-minute and one-hour volatility estimates.

Consistent with previous studies we show in Table 5 Panel A, that the one-hour volatility on nonfarm payroll announcement days is 3 to 6 times larger than during non-announcement days. The one-minute volatility is 15 to 30 times larger during announcement days compared to non-announcement days. As expected given the fact that we focus on a U.S. data release, the volatility increase is smaller in the cross-rate, the euro-yen exchange rate, than in the euro-dollar and yen-dollar exchange rates. Focusing on the statistically significant estimates, we show in Table 5 Panel B that, as a share of total volume, human-maker volume tends to increase during the minute of the announcement (the one-minute ratio $\frac{LH_a}{LH_n}$ is greater than one), while computer-maker volume tends to decrease (the one-minute ratio $\frac{LC_a}{LC_n}$ is less than one). Interestingly, this pattern is reversed when we focus on the one-hour volume estimates for the euro-dollar and euro-yen exchange rate markets. In relative terms, computers do not increase their provision of liquidity as much as humans do during the minute following the announcement. However, computers increase their provision of liquidity relatively more than humans do over the entire hour following the announcement, a period when market volatility remains quite elevated.

We note that, over our sample period, the U.S. nonfarm payroll data releases were clearly the most anticipated and most influential U.S. macroeconomic data releases. They often generated a large initial sharp movement in exchange rates, followed by an extended period of volatility. The behavior of computer traders observed in the first minute could reflect the fact that many algorithms are not designed to react to sharp, almost discrete, moves in exchange rates. Some algorithmic traders may then prefer to pull back from the market a few seconds before 8:30 a.m. ET on days of nonfarm payroll announcements, resuming trading once the risk of a sharp initial price movement has passed. But the data show that algorithmic traders, as a whole, do not shrink back from providing liquidity during the extended period of volatility that follows the data releases.

[16] For simplicity, we label the 10 business days before and after the nonfarm payroll announcement as non-announcement days. However, during this 20-day period there are days with no macroeconomic news and days with news, e.g., every Thursday, the day before the nonfarm payroll number is released, initial claims are released. Thus our estimation will be biased towards not finding statistically different behavior across the two periods. As we show in Table 5, volatility is, on average, much lower during this 20-day period and thus serves as a good non-announcement benchmark period. We note, though, that our results are stronger when we drop announcement days from the 20 days surrounding nonfarm payroll release dates and consider that as the non-announcement benchmark period.

6 Price Discovery

In the previous three sections, we analyze questions that are primarily motivated by practical concerns regarding algorithmic trading, such as whether computer traders induce volatility or reduce liquidity. In this section we turn to questions that are more driven by the market microstructure literature, but that also lead to interesting practical insights regarding the effects and nature of algorithmic trading. In particular, we study price discovery within a vector autoregressive framework, which enables us to evaluate to what extent humans or computers represent the "informed" traders in the market. Our findings reveal several interesting features regarding the impact of algorithmic trades and the order placement behavior of computer traders.

6.1 Who is the "informed" trader, humans or computers?

In this section we investigate whose trades, human's or computer's, have a permanent impact on prices. To this end, we estimate return-order flow dynamics in a structural vector autoregressive (VAR) framework in the tradition of Hasbrouck (1991a), where returns are contemporaneously affected by order flow, but order flow is not contemporaneously affected by returns. Similar to Hasbrouck's (1996) decomposition of program and nonprogram order flow, we decompose order flow into two components: human-taker $\left(OF^{(ht)} = HH + CH\right)$ and computer-taker $\left(OF^{(ct)} = HC + CC\right)$, and thus we estimate for each currency i one return equation and two order flow equations. In light of Evans and Lyons (2008) findings, we estimate the structural VAR with U.S. macroeconomic news surprises as exogenous variables that affect both returns and order flow. Specifically, we estimate the following system of equations for each currency i,

$$\begin{aligned}
r_{it} &= \alpha^r + \sum_{j=1}^{J}\beta_{ij}^{r} r_{it-j} + \sum_{j=0}^{J}\gamma_{ij}^{rct} OF_{it-j}^{(ct)} + \sum_{j=0}^{J}\gamma_{ij}^{rht} OF_{it-j}^{(ht)} + \sum_{k=1}^{K}\delta_{ik}^{r} S_{kt} + \varepsilon_{it}^{r}, \qquad (2)\\
OF_{it}^{(ht)} &= \alpha_{ht}^{OF} + \sum_{j=1}^{J}\beta_{ijht}^{OF} r_{it-j} + \sum_{j=1}^{J}\gamma_{ijht}^{OF^{(ht)}} OF_{it-j}^{(ht)} + \sum_{j=1}^{J}\gamma_{ijht}^{OF^{(ct)}} OF_{it-j}^{(ct)} + \sum_{k=1}^{K}\delta_{ikht}^{OF} S_{kt} + \varepsilon_{it}^{OF^{(ht)}},\\
OF_{it}^{(ct)} &= \alpha_{ct}^{OF} + \sum_{j=1}^{J}\beta_{ijct}^{OF} r_{it-j} + \sum_{j=1}^{J}\gamma_{ijct}^{OF^{(ct)}} OF_{it-j}^{(ct)} + \sum_{j=1}^{J}\gamma_{ijct}^{OF^{(ht)}} OF_{it-j}^{(ht)} + \sum_{k=1}^{K}\delta_{ikct}^{OF} S_{kt} + \varepsilon_{it}^{OF^{(ct)}}.
\end{aligned}$$

Here r_{it} is the 1-minute exchange rate return for currency i at time t; OF_{it}^{ht} is the currency i human-taker order flow at time t; OF_{it}^{ct} is the currency i computer-taker order flow at time t; and S_{kt} is the macroeconomic news announcement surprise for announcement k at time t defined as the difference between the announcement realization and its corresponding market expectation. We use Bloomberg's real-time data on the expectations and realizations of $K = 28$ U.S. macroeconomic fundamentals to calculate S_{kt}. The 28 announcements we consider are similar to those in Andersen et al. (2003, 2007) and Pasquariello and Vega (2007).[17] Since units of measurement vary across macroeconomic variables, we standardize the resulting surprises by dividing each of them by their sample standard deviation. Economic theory suggests that we should also include foreign macroeconomic news announcements in equation (2). However, previous studies find that exchange rates do not respond much to non-U.S. macroeconomic announcements, even at high frequencies, e.g. Andersen et al. (2003), so we expect the omitted variable bias in our specification to be small.

The underlying economic model is based in continuous time, and we thus estimate the VAR using the highest sample frequency available to us, minute-by-minute data.[18] The estimation period is restricted to the $2006-2007$ sample, and the total number of observations for each currency pair is $717,120$ in the full sample and $89,280$ in the three month sub-sample (September, October and November of 2007). In both samples, 20 lags are included in the estimated VARs, i.e. $J = 20$.

Before considering the impulse response functions and the variance decompositions, we briefly summarize the main lessons from the estimated coefficients in the VAR. Focusing on the return equation, we find that minute-by-minute returns tend to be negatively serially correlated, with the coefficient on the first own lag varying between -0.08 and -0.15; there is thus some evidence of mean reversion in the exchange rates at these high frequencies, which is a well-know empirical finding. Both order flows are significant predictors of returns. The price impact of the lagged order flows range from around 4 to 18 basis points per billion units of order flow (denominated in the base currency), as compared to a range of approximately $28-100$ basis points in the contemporaneous order flow. As theory would predict, we find that U.S. macroeconomic news announcements affect less the euro-yen exchange rate (i.e., the R^2 of regressing the euro-yen exchange rate on macroeconomic news surprises and restricting the sample to announcement-only observations is 23%) than the euro-dollar and dollar-yen exchange rates (i.e., the R^2 of an announcement-only sample is 60% and 59%, respectively). However, U.S. macroeconomic news announcements still have an effect on the cross-rate to the extent that the U.S. economy is more or less correlated with the Japanese or the Euro-area economy.

Focusing on the order-flow equations, we find that the first own lag in both order flow equations is always highly significant, and typically around 0.1 for all currency pairs. There is thus a sizeable first-order autocorrelation in the human-taker and computer-taker order flows. The coefficients on the first order cross-

[17] Our list of U.S. macroeconomic news announcements is the same as the list of announcements in Andersen et al. (2007) and Pasquariello and Vega (2007) with the addition of three announcements: unemployment report, core PPI and core CPI. Andersen et al. (2007) and Pasquariello and Vega (2007) use International Money Market Services (MMS) data on the expectations of U.S. macroeconomic fundamentals, in contrast, we use Bloomberg data because MMS was acquired by Informa Global in 2003 and no longer exists. Bloomberg, though, provides similar survey data to those MMS has provided.

[18] At the one-minute frequency, as opposed to in a tick-by-tick setting, the possibility of simultaneity among the variables does exist. For example, a quote revision that arrives to the market at the beginning of a given minute could conceivably influence the order flow arriving at the end of that minute. While we acknowledge this possibility, we do not believe that it is of serious concern at the very high frequency of our analysis. Daníelsson and Love (2004) explicitly try to estimate the effects of such feedback trading in foreign exchange data. They find that feedback effects can play a significant role in data sampled at the five-minute frequency, but less so in data sampled at the one-minute frequency.

lags in the order flow regressions are most often substantially smaller than the coefficient on the own lag and vary in signs. Lagged returns have a small but positive impact on order flow, suggestive of a form of 'trend chasing' by both computers and humans in their order placement.

We note that despite the strongly significant estimates that are recorded in the VAR estimations, the amount of variation in the order flow and return variables that is captured by their lagged values is very limited. The R^2 for the estimated equations with only lagged variables are typically around three to ten percent for the order flow equations, and between one and three percent for the return equations. Compared to an R^2 of 20 to 30 percent when one includes contemporaneous order flow.

6.2 Impulse Response Function and Variance Decomposition Results

As originally suggested by Hasbrouck (1991b), we use the impulse response functions to assess the price impact of various order flow types, and the variance decompositions to measure the relative importance of the variables driving foreign exchange returns. In Table 6 Panel A, we show the results from the impulse response analysis based on the estimation of equation (2), using the full sample for 2006-2007 and the three-month sub-sample, when the size of the shock is the same across the different types of order flow: a one billion base currency shock to order flow. We also show the results when the size of the shock varies according to the average size shock: a one standard deviation base currency shock to order flow (Table 6 Panel B). We show both the short-run (instantaneous) impulse responses, the long-run cumulative responses, and the difference between the two responses. The long-run statistics are calculated after 30-minutes, at which point the cumulative impulse responses have converged and can thus be interpreted as the long-run total impact of the shock. All the responses are measured in basis points.

Starting with a hypothetical shock of one billion base currency order flow, the results in Table 6 Panel A, show that the immediate response of prices to human-taker order flow is often larger than the immediate response to computer-taker order flow. This may partially be attributed to the fact that some of the algorithmic trading is used for the optimal execution of large orders at a minimum cost. Algorithmic trades appear to be successful in that endeavor, with computers breaking up the larger orders and having a minimum impact on prices. We emphasize that the differences in price impact, although statistically significant, range from 1 to 8 basis points and are thus not that large in economic terms. Furthermore, we often find that the result is reversed in the long-run and in the three-month sub-sample. For example, the euro-dollar human-taker price impact is larger than the computer-taker price impact in the short-run, but the opposite is true in the long-run and in the three month sub-sample.

In contrast to these results, the response to a hypothetical one standard deviation shock to the different order flows (Table 6 Panel B) consistently shows that in the euro-dollar and dollar-yen exchange rate markets, humans have a bigger impact on prices than computers and the differences are relatively large. For example, a one standard deviation shock to human-taker order flow in the yen-dollar exchange rate market has an average long-run effect of 0.9 basis points compared to an average effect of 0.3 basis points for computer-taker order flow. Interestingly, the difference in price impact in the cross-rate, the euro-yen exchange rate, is very small. In this market, computers have a clear advantage over humans in detecting and reacting more quickly to triangular arbitrage opportunities so that a large proportion of algorithmic trading contributes to more efficient price discovery. It is then not so surprising that in this market computers and humans, on average, appear to be equally "informed." This finding suggests that the effect order flow has on prices may not so much be a matter of whether computers or humans are trading, but a matter of what computers are predominantly used for.

In Table 7 we report what fraction of the total (long-run) variance in returns that can be attributed to innovations in human-taker order flow and computer-taker order flow.[19] Following Hasbrouck (1991b), we interpret this variance decomposition as a summary measure of the informativeness of trades, and thus, in the current context, a comparison of the relative informativeness of the different types of order flow. Consistent with the results from the impulse response functions based on a one standard deviation shock to order flow, we find that in the euro-dollar and dollar-yen exchange rate markets human-taker order flow explains more of the total variance in returns than computer-taker order flow. Specifically, human-taker order flow explains about 30 percent of the total variance in returns compared to only 4 percent explained by computer-taker order flow. The fact that human-taker order flow explains a bigger portion of total variance in returns may not be surprising because human-taker volume is about 75 percent of total volume in these two markets in the full sample period and about 65 percent of total volume in the three-month sub-sample (see Table 2). Moreover large buy (sell) orders tend to be human-taker orders, i.e. we show in Table 1 that the standard deviation of human-taker order flow is twice as big as that of the computer-taker order flow.

Do computers tend to contribute 'disproportionately' little to the long-run variance in returns relative to their trading volume? To answer this question we do a back-of-the-envelope calculation. We compute the relative share of the explained variance that is due to computer-taker order flow as the percent of total variation in returns explained by computer-taker order flow divided by the percent of total variation in returns explained by total order flow (human-taker plus computer-taker order flow). For example, this relative share is $14\% = 100 \times \frac{4.74}{34}$ (Table 7) in the euro-dollar market. We can then compare this relative share to the fraction of overall trading volume that is due to computer-taker volume, which we show in Table 2. In the

[19] The variance decompositions are virtually identical in the short- and long-run and thus we only show the long-run decomposition results.

full 2006-2007 sample for the euro-dollar and the dollar-yen currency pairs, the fraction of volume due to computer-takers is about twice as large as the fraction of the explained long-run variance that is due to computer-taker order flow. In the euro-yen, the fractions are approximately equal. These results are fairly similar in the three-month sub-sample, although the fraction of explained variance has increased somewhat compared to the volume fraction. Thus, in the two major currency pairs, there is evidence that computer-taker order flow contributes relatively less to the variation in returns than one would infer from just looking at the proportion of computer-taker volume.

The seemingly disproportionately small fraction of the explained return variance that can be attributed to computer-taker order flow is a result both of the generally smaller responses by returns to a one-billion base currency shocks from this order flow component (Table 6 Panel A), as well as the generally smaller shocks that occur in this order flow as seen from the estimates of the standard deviation in the different order flows (Table 1).[20,21] However, the results in Table 6 Panel A suggest that it is more due to the latter than to the former. This makes economic sense because human-makers and computer-makers can not identify ex ante trades coming from computers or humans. Thus, fixing the order size, computer and human trades should have similar impact on prices.

6.3 Are liquidity providers "uninformed"?

We now turn to examine whether liquidity providers strategically post quotes. To this end we augment equation (2) and decompose order flow into four components. Specifically, we estimate the following system of equations for each currency i,

$$\begin{aligned} r_{it} &= \alpha^r + \sum_{j=1}^{J} \beta_{ij}^r r_{it-j} + \sum_{l=1}^{L}\sum_{j=0}^{J} \gamma_{ij}^{rl} OF_{it-j}^{(l)} + \sum_{k=1}^{K} \delta_{ik}^r S_{kt} + \varepsilon_{it}^r, \\ OF_{it}^{(l)} &= \alpha_l^{OF} + \sum_{j=1}^{J} \beta_{ijl}^{OF} r_{it-j} + \sum_{l=1}^{L}\sum_{j=1}^{J} \gamma_{ijl}^{OF^{(l)}} OF_{it-j}^{(l)} + \sum_{k=1}^{K} \delta_{ikl}^{OF} S_{kt} + \varepsilon_{it}^{OF^{(l)}}. \end{aligned} \tag{3}$$

where r_{it} is the 1-minute exchange rate return for currency i at time t; $L = 4$, $OF_{it}^{(1)} = OF_{it}^{HH}$ is the currency i human-maker/human-taker order flow at time t; $OF_{it}^{(2)} = OF_{it}^{CH}$ is the currency i computer-maker/human-taker order flow at time t; $OF_{it}^{(3)} = OF_{it}^{HC}$ is the currency i human-maker/computer-taker order flow at time t; $OF_{it}^{(4)} = OF_{it}^{CC}$ is the currency i computer-maker/computer-taker order flow at time t;

[20] The variance decompostion is a function of the (squared) terms in the Vector Moving Average (VMA) representation of the VAR and the variance of the shocks in the VAR equations (i.e. the variance of the VAR residuals). For a given shock size, the impulse response functions are a function of the (non-squared) VMA coefficients.

[21] Strictly speaking, the variance decomposition is a function of the variance in the shocks in the VAR residuals and not in the original data entering the VAR, i.e. the variance of the unexpected shocks. However, since the R^2s in the VAR equations are small, the variance in the VAR residuals and the original data are very similar.

S_{kt} is the macroeconomic news announcement surprise for announcement k at time t.

In addition to identifying whether traders, on average, have a more permanent impact on prices when trading with humans than with computers, this specification also allows us to observe the effect order flow has on prices when, for instance, no party has a speed advantage, i.e. both parties are humans or both parties are computers, and when either the maker has a speed advantage, CH, or the taker has a speed advantage, HC. This distinction may be particularly useful when analyzing the cross-rate, where computers likely have a clear advantage over humans in detecting short-lived triangular arbitrage opportunities.

Starting with a hypothetical shock of one billion base currency order flow, the results in Table 8 Panel A show that there is no clear pattern in whose order flow impacts prices the most. The dynamics of the VAR system help reveal an interesting finding, though: There is a consistent and often large short-run over-reaction to CC and CH shocks. That is, as seen in Table 8 the short run response to a CC or CH order flow shock is always larger than the long-run response, and sometimes substantially so. The euro-dollar in the sample covering September, October, and November of 2007 provides an extreme case where the initial reaction to a one billion dollar CC shock is a 22 basis point move, but the long-run cumulative reaction is just 6 basis points. Interestingly, the opposite pattern is true for the HH order flow shocks, where there is always an initial *under*-reaction in returns. To the extent that an over-reaction of prices to order flow is suggestive of the presence of liquidity traders, these impulse response patterns suggest that computers provide liquidity when the probability of trading with an informed trader is low.[22]

The response to a hypothetical one standard deviation shock to the different order flows consistently shows that HH order flow has a bigger impact on prices than CC order flow (Table 8 Panel B) and that the differences are large. In particular, a one standard deviation shock to HH order flow has an average long-run effect of 0.6 basis points across currencies compared to a one standard deviation shock to CC order flow, which has an average effect of 0.1 basis points. Interestingly, we observe that when humans trade with other humans they influence prices more than when they trade with computers (the impact of HH on prices is bigger than the impact of CH on prices), and when computers trade with other computers they influence prices less than when they trade with humans (the impact of CC on prices is bigger than the impact of HC on prices). Our interpretation is that computers provide liquidity more strategically than humans, so that the counterparty cannot affect prices as much. This interpretation is consistent with the over-reaction of prices to CC and CH order flow described above: Computers appear to provide liquidity when adverse selection costs are low. This finding relates to the literature that proposes to depart from the prevalent assumption

[22] Dynamic learning models with informed and uninformed investors predict that prices will temporarily over-react to uninformed order flow and under-react to informed order flow (e.g., Albuquerque and Miao, 2008). We note that the over- and under-reaction of prices to a particular type of order flow is different from the over- and under-reaction of prices to public news, which are both considered a sign of market inefficiency. Order flow types are not public knowledge, so that agents cannot trade on this information.

that liquidity providers in limit order books are passive.[23]

We also find that the price response to order flow varies across currencies as these markets differ along several dimensions. Trading volume is largest in the euro-dollar and dollar-yen markets, compared to the euro-yen market, and price discovery clearly happens mostly in the two largest markets. In the cross-rate market, the euro-yen, computers have a speed advantage over humans in profiting from triangular arbitrage opportunities, where prices set in the euro-dollar and dollar-yen markets are very briefly out of line with the euro-yen rate. Consistent with this speed advantage we observe that human-maker/computer-taker order flow has a larger price impact in the cross-rate market than in the other two markets.

In addition to the impulse response functions, we also report the long-run forecast variance decomposition of returns in Table 9 for both the full sample and the three-month sub-sample. Consistent with the impulse response functions to a one standard deviation shock to order flow, the HH order flow makes up the dominant part of the variance share in the euro-dollar and dollar-yen exchange rate markets. In the last three months of the sample, this share has generally decreased. The share of variance in returns that can be attributed to the CC order flow is surprisingly small, especially in the latter sub-sample, given that this category of trades represent a sizeable fraction of overall volume of trade during the last months of 2007, as seen in Table 2. The mixed order flow (CH and HC order flow) typically contributes with about the same share to the explained variance in the euro-dollar and dollar-yen exchange rate markets. In contrast, in the euro-yen exchange rate market HC order flow makes up the dominant part of the variance share, which is consistent with our discussion of computers taking advantage of triangular arbitrage opportunities in this market.

Overall, about 15 to 35 percent of the total variation in returns can be attributed to shocks to the four order flows. However, in most currency pairs, very little of this ultimate long-run price discovery that occurs via order flow does so via the CC order flow. Similar to Table 7, we also report in Table 9 the fraction of the explained share of the return variance that can be attributed to the different order flow combinations. Comparing these to the fraction of overall volume that is due to these combinations of computers and humans, reported in Table 2, gives an idea of whether the different order flow combinations contribute proportionately to the variance in returns. It is clear that CC order flow tends to contribute disproportionately little to the long-run variance of returns, and that HH order flow often contributes disproportionately much.

[23] For example, Chakravarty and Holden (1995), Kumar and Seppi (1994), Kaniel and Liu (2006), and Goettler, Parlour and Rajan (2007) allow informed investors to use both limit and market orders. Bloomfield, O'Hara and Saar (2005) argue that informed traders are natural liquidity providers and Angel (1994) and Harris (1998) show that informed investors can optimally use limit orders when private information is sufficiently persistent.

7 Conclusion

Using high-frequency trading data for three exchange rates over 2006 and 2007, we analyze the impact of the growth of algorithmic trading on the spot interdealer foreign exchange market. In particular, we try to answer the following questions. (i) Are the algorithms underlying the computer generated trades similar enough to result in highly correlated strategies, which may cause disruptive market behavior? (ii) Does algorithmic trading increase volatility in the market, perhaps as a result of the previous concern? (iii) Do algorithmic traders improve or reduce liquidity at times when it is needed the most? (iv) Are human or computer traders the more "informed" traders in the market, i.e. who has the most impact on price discovery? (v) Is there evidence in this market that liquidity providers (the "makers") and not just liquidity "takers", are informed, and do computer makers post orders more strategically than humans?

The first three of these questions are primarily motivated by concerns that have been raised in the popular press, especially in conjunction with the current financial crisis, whereas the last two questions obviously relate to the empirical market microstructure literature on price discovery and order placement. However, the analysis of all five questions brings new interesting results to the table, both from a practical and academic perspective.

Our empirical results show that there is a tendency for algorithmic trades to be more correlated than non-algorithmic trades, suggesting that the trading strategies used by the automated computer traders are less diverse than those of their human counterparts. Although this may cause some concerns regarding the disruptive potential of computer generated trades, we do not find any evidence of a positive causal relationship between the proportion of algorithmic trading in the market and the level of volatility; if anything, the evidence points towards a negative relationship, suggesting that the presence of algorithmic trading reduces volatility. As for the provision of market liquidity by the different types of traders, we find evidence that, compared to non-algorithmic traders, algorithmic traders reduce their share of liquidity provision in the minute following major data announcements, when the probability of a price jump is very high. However they increase their share of liquidity provision to the market over the entire hour following the announcement, which is almost always a period of elevated volatility. This empirical evidence thus suggests that computers *do* provide liquidity during periods of market stress. Overall, there is little statistical evidence to support popular concerns that algorithmic trading increases volatility or decreases liquidity.

To address questions (iv) and (v) above, we use a high-frequency VAR framework in the tradition of Hasbrouck (1991a). We find that non-algorithmic trades account for a substantially larger share of the price movements in the euro-dollar and yen-dollar exchange rate markets than would be expected given the sizable fraction of algorithmic trades, i.e., non-algorithmic traders are the "informed" traders in these markets. In the cross-rate, the euro-yen exchange rate market, we find that computers and humans are equally "informed"; coincidentally, we believe that in this market a large proportion of algorithmic trades help make markets more efficient by taking advantage of triangular arbitrage opportunities. This finding suggests that, in the analysis of the effect that order flow from different sources has on prices, focusing on what algorithmic trading is predominantly used for may be as important as whether or not the order flow is generated by computers. In addition, we find that, on average, computer-takers or human-takers that trade with a computer-maker do not impact prices as much as they do when they trade with a human-maker. One interpretation of this result is that computers place limit orders more strategically than humans do. This finding dovetails with the literature on limit order books that relaxes the common assumption that liquidity providers are passive.

References

[1] Albuquerque, R., and Miao, J., 2008, Advanced Information and Asset Prices, Working Paper Boston University.

[2] Andersen, T.G., and T. Bollerslev, 1998, Deutsche Mark-Dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies, *Journal of Finance* 53, 219-265.

[3] Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens, 2001. The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43–76.

[4] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2003. Modeling and Forecasting Realized Volatility, *Econometrica* 71, 579-625.

[5] Andersen, T.G., T. Bollerslev, F.X. Diebold, and C. Vega, 2003. Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange, *American Economic Review* 93, 38-62.

[6] Andersen, T.G., T. Bollerslev, F.X. Diebold, and C. Vega, 2007. Real-time price discovery in global stock, bond, and foreign exchange markets, *Journal of International Economics* 73, 251-277.

[7] Angel, J., 1992, Limit versus Market Orders, working paper, Georgetown University.

[8] Bandi, F.M., and J.R. Russel, 2006. Separating microstructure noise from volatility, *Journal of Financial Economics* 79, 655-692.

[9] Berger, D.W., A.P. Chaboud, S.V. Chernenko, E. Howorka, and J.H. Wright, 2008. Order Flow and Exchange Rate Dynamics in Electronic Brokerage System Data, *Journal of International Economics* 75, 93-109.

[10] Biais, B., Hillion, P., and Spatt, C., 1995, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *Journal of Finance*, 50, 1655-1689.

[11] Bollerslev T., and J.H. Wright, 2000. Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data, *Journal of Econometrics* 98, 81-106.

[12] Bloomfield, R., M. O'Hara and G. Saar, 2005, The "Make or Take" Decision in an Electronic Market: Evidence on the Evolution of Liquidity, *Journal of Financial Economics* 75, 165–199.

[13] Chaboud, A., B. Chiquoine, E. Hjalmarsson, and M. Loretan, 2007. Frequency of Observation and the Estimation of Integrated Volatility in Deep and Liquid Markets, International Finance Discussion Paper 905, Board of Governors of the Federal Reserve Board.

[14] Chakravarty, S., and C. Holden, 1995, An Integrated Model of Market and Limit Orders, *Journal of Financial Intermediation* 4, 213–241.

[15] Daníelsson, J., and R. Love, 2004. Feedback Trading, Working Paper, London School of Economics.

[16] Easley, D., and O'Hara, M., 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics*, 19, 69-90.

[17] Evans, M., and R. Lyons, 2002. Order Flow and Exchange Rate Dynamics, *Journal of Political Economy* 110, 170-180.

[18] Evans, M., and R. Lyons, 2008. How is Macro News Transmitted to Exchange Rates?, *Journal of Financial Economics* 88, 26-50.

[19] Goettler, R., C. Parlour and U. Rajan, 2007, Microstructure effects and Asset pricing, working paper, UC Berkeley.

[20] Harris, L., 1998, Optimal Dynamic Order Submission Strategies in Some Stylized Trading Problems, *Financial Markets, Institutions & Instruments* 7, 1–76.

[21] Hasbrouck, J., 1991a. Measuring the information content of stock trades, *Journal of Finance* 46, 179-207.

[22] Hasbrouck, J., 1991b. The summary informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 4, 571-595.

[23] Hasbrouck, J., 1996. Order characteristics and stock price evolution: An application to program trading, *Journal of Financial Economics* 41, 129-149.

[24] He, H., and J. Wang, 1995, Differential information and dynamic behavior of stock trading volume, *Review of Financial Studies*, 8, 919-972.

[25] Kaniel, R., and H. Liu, 2006, What Orders do Informed Traders Used, *Journal of Business* 79, 1867–1913.

[26] Kumar, P., and D. Seppi, 1994, Limit and Market Orders with Optimizing Traders, working paper, Carnegie Mellon University.

[27] Madhavan, A., Richardson, M., and Roomans, M., 1997, Why do security prices change? A transaction-level analysis of NYSE stocks, *Review of Financial Studies*, 10, 1035-1064.

[28] Milgrom, P., and Stokey, N., 1982, Information, trade and common knowledge, *Journal of Economic Theory*, 26, 17-27.

[29] Parlour, Christine, and Seppi, Duane, 2008, Limit Order Markets: A Survey, *Handbook of Financial Intermediation & Banking*, ed. A.W.A. Boot and A.V. Thakor, forthcoming.

[30] Pasquariello, Paolo, and Vega, Clara, 2007, Informed and Strategic Order Flow in the Bond Markets, *Review of Financial Studies*, 20, 1975-2019.

[31] Stock, J.H., J.H. Wright, and M. Yogo, 2005. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics*, 20, 518-529.

[32] Stock, J.H., and M. Yogo, 2005. Testing for Weak Instruments in Linear IV Regression, in D.W.K. Andrews and J.H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge: Cambridge University Press, 80–108.

Table 1: Summary statistics for the one-minute return and order flow data. The mean and standard deviation, as well as the first-order autocorrelation, ρ, are shown for each variable and currency pair. The returns are expressed in basis points and the order flows in millions of the base currency. The summary statistics are given for both the full 2006-2007 sample, as well as for the three-month sub-sample, which only uses observations from September, October, and November of 2007. The first two rows for each currency shows the summary statistics for returns and the total market-wide order flow. The following two rows give the results for the order flows broken down into human takers and computer takers and the last four rows show the results for the order flow decomposed into each maker-taker pair. There are a total of $717,120$ observations in the full two year sample and $89,280$ observations in the three month sub sample. We show the statistical significance of the first order autocorrelation. The ***, **, and * represents significance at the 1, 5, and 10 percent level, respectively.

	Full 2006-2007 Sample			3-month sub sample		
Variable	Mean	Std. dev.	ρ	Mean	Std. dev.	ρ
			USD/EUR			
Returns	0.0030	1.2398	-0.005^{***}	0.0080	1.2057	0.007^{**}
Total order flow $(HH+CH+HC+CC)$	0.0315	25.9455	0.150^{***}	-0.0937	29.7065	0.174^{***}
H-taker $(HH+CH)$	0.0413	23.977	0.155^{***}	-0.0796	26.8096	0.189^{***}
C-taker $(HC+CC)$	-0.0099	9.9363	0.127^{***}	-0.0140	12.8900	0.115^{***}
H-maker/H-taker (HH)	0.1425	19.9614	0.177^{***}	0.0327	21.9211	0.209^{***}
C-maker/H-taker (CH)	-0.1012	8.8970	0.166^{***}	-0.1123	10.7649	0.215^{***}
H-maker/C-taker (HC)	0.0123	8.9232	0.152^{***}	0.0483	11.5856	0.150^{***}
C-maker/C-taker (CC)	-0.0222	2.7939	0.053^{***}	-0.0623	3.9477	0.072^{***}
			JPY/USD			
Returns	-0.0007	1.6038	-0.010^{***}	-0.0045	1.9110	0.007^{**}
Total order flow $(HH+CH+HC+CC)$	0.1061	20.0980	0.189^{***}	-0.3439	23.6359	0.211^{***}
H-taker $(HH+CH)$	0.0853	19.1127	0.190^{***}	-0.2088	22.0344	0.204^{***}
C-taker $(HC+CC)$	0.0209	8.3941	0.170^{***}	-0.1351	11.5877	0.158^{***}
H-maker/H-taker (HH)	0.1037	15.9972	0.209^{***}	-0.1203	17.4612	0.226^{***}
C-maker/H-taker (CH)	-0.0184	6.9030	0.172^{***}	-0.0885	9.1773	0.162^{***}
H-maker/C-taker (HC)	0.0198	7.5686	0.198^{***}	-0.0901	10.1673	0.191^{***}
C-maker/C-taker (CC)	0.0011	2.4556	0.032^{***}	-0.045	3.8751	0.026^{***}
			JPY/EUR			
Returns	0.0024	1.5976	-0.053^{***}	0.0036	2.1398	-0.017^{***}
Total order flow $(HH+CH+HC+CC)$	-0.0648	7.0941	0.152^{***}	-0.1574	8.5978	0.147^{***}
H-taker $(HH+CH)$	-0.0497	5.7006	0.150^{***}	-0.1216	6.2074	0.125^{***}
C-taker $(HC+CC)$	-0.0151	4.8409	0.146^{***}	-0.0358	6.7000	0.131^{***}
H-maker/H-taker (HH)	-0.0172	4.4203	0.159^{***}	-0.0600	4.3106	0.157^{***}
C-maker/H-taker (CH)	-0.0325	2.8912	0.129^{***}	-0.0616	3.7197	0.092^{***}
H-maker/C-taker (HC)	-0.0095	4.5331	0.173^{***}	-0.0264	6.0968	0.161^{***}
C-maker/C-taker (CC)	-0.0056	1.5558	0.023^{***}	-0.0095	2.5621	-0.001

Table 2: Summary statistics for the fractions of trade volume attributable to different trader combinations. The table shows the fraction of the total volume of trade that are attributable to different combinations of makers and takers. Results for the full 2006-2007 sample as well as for the three-month sub-sample, which only uses data from September, October, and November of 2007, are shown. We show the average of the daily fractions, calculated by summing up across all minutes within a day, and the standard deviations of those daily fractions. For each currency, the first row shows the fraction of the total volume of trade where a computer was involved on at least one side of the trade (i.e. as a maker or a taker). The second row shows the fraction of the total volume where a human acted as a taker, the third row shows the fraction of the total volume where a computer acted as a taker, and the following four rows shows the fractions broken down by each maker-taker pair.

	Full 2006-2007 Sample		3-month sub sample	
Variable	mean	Std. dev.	mean	Std. dev.
	USD/EUR			
C-participation ($Vol\,(CH + HC + CC)$)	0.4163	0.1135	0.5386	0.0355
H-taker ($Vol\,(CH + HH)$)	0.7810	0.0791	0.6864	0.0331
C-taker ($Vol\,(HC + CC)$)	0.2190	0.0791	0.3136	0.0331
H-maker/H-taker ($Vol\,(HH)$)	0.5837	0.1135	0.4614	0.0355
C-maker/H-taker ($Vol\,(CH)$)	0.1973	0.0398	0.2251	0.0144
H-maker/C-taker ($Vol\,(HC)$)	0.1710	0.0514	0.2304	0.0205
C-maker/C-taker ($Vol\,(CC)$)	0.0480	0.0290	0.0831	0.0150
	JPY/USD			
C-participation ($Vol\,(CH + HC + CC)$)	0.4242	0.1065	0.5652	0.0364
H-taker ($Vol\,(CH + HH)$)	0.7585	0.0805	0.6461	0.0311
C-taker ($Vol\,(HC + CC)$)	0.2415	0.0805	0.3539	0.0311
H-maker/H-taker ($Vol\,(HH)$)	0.5758	0.1065	0.4348	0.0364
C-maker/H-taker ($Vol\,(CH)$)	0.1827	0.0304	0.2114	0.0126
H-maker/C-taker ($Vol\,(HC)$)	0.1860	0.0498	0.2486	0.0154
C-maker/C-taker ($Vol\,(CC)$)	0.0555	0.0321	0.1052	0.0193
	JPY/EUR			
C-involved ($Vol\,(CH + HC + CC)$)	0.6186	0.1154	0.7907	0.0410
H-taker ($Vol\,(CH + HH)$)	0.5557	0.1018	0.4037	0.0467
C-taker ($Vol\,(HC + CC)$)	0.4443	0.1018	0.5963	0.0467
H-maker/H-taker ($Vol\,(HH)$)	0.3814	0.1154	0.2093	0.0410
C-maker/H-taker ($Vol\,(CH)$)	0.1743	0.0360	0.1944	0.0164
H-maker/C-taker ($Vol\,(HC)$)	0.3337	0.0473	0.3734	0.0193
C-maker/C-taker ($Vol\,(CC)$)	0.1106	0.0673	0.2229	0.0464

Table 3: Estimates of the ratio $R = RC/RH$. The table reports, for various sub-samples, the mean estimates of the ratio $R = RC/RH$, where $RC = Vol(HC)/Vol(CC)$ and $RH = Vol(HH)/Vol(CH)$. $Vol(HH)$ is the daily trading volume between human makers and human takers, $Vol(HC)$ is the daily trading volume between human makers and computer takers, $Vol(CH)$ is the daily trading volume between computer makers and human takers, and $Vol(CC)$ is the daily trading volume between computer makers and computer takers. We report the mean of the daily ratio R and the standard errors are shown in parantheses below the estimate. We also show the number of days that had a ratio that was less than one. We report the results for the full 2006-2007 sample and the three-month sub-sample, which only uses data from September, October, and November of 2007. The ***, **, and * represents a statistically significant deviation from one at the 1, 5, and 10 percent level, respectively.

	Full 2006-2007 sample	3-month sub sample
	USD/EUR	
Mean of daily $R = RC/RH$	1.4463***	1.3721***
Standard Error	(0.0063)	(0.0099)
No. of days with $R < 1$	0	0
No. of obs	498	62
	JPY/USD	
Mean of daily $R = RC/RH$	1.2619***	1.1719***
Standard Error	(0.0074)	(0.0142)
No. of days with $R < 1$	15	4
No. of obs	498	62
	JPY/EUR	
Mean of daily $R = RC/RH$	1.6886**	1.6242***
Standard Error	(0.0154)	(0.0250)
No. of days with $R < 1$	4	0
No. of obs	498	62

Table 4: Regressions of realized volatility on the fraction of algorithmic trading. The table shows the results from estimating the relationship between daily realized volatility and the fraction of algorithmic trading, using daily data from 2006 and 2007. Robust standard errors are given in parentheses below the coefficient estimates. The left hand side of the table shows the results with a quarterly time trend included in the regressions and the right hand side of the table shows the results with year-quarter time dummies (i.e., eight time dummies, one for each quarter in the two years of data) included in the regressions. Panels A and B show the results when the fraction of algorithmic trading is measured as the fraction of the total trade volume that has a computer involved on at least one side of the trade (i.e. as a maker or a taker). Panels C and D show the results when only the fraction of volume with computer taking is used. In addition to the fraction of algorithmic trading and the control(s) for secular trends, 22 lags of volatility are also included in every specification. In all cases, only the coefficient on the fraction of algorithmic trading is displayed. Panels A and C show the results from a standard OLS estimation, along with the R^2. Panels B and D show the results from the IV specification estimated with Limited Information Maximum Likelihood (LIML). In Panels B and D, the Stock and Yogo (2005) $F-$test of weak instruments are also shown. The critical values for Stock and Yogo's (2005) F-test are designed such that they indicate a maximal actual size for a nominal sized five percent test on the coefficient in the LIML estimation. Thus, in order for the actual size of the LIML test to be no greater than 10% (15%), the F-statistic should exceed 8.68 (5.33). There are a total of 498 daily observations in the data. The ***, **, and * represents significance at the 1, 5, and 10 percent level, respectively.

	With quarterly time trend			With year-quarter time dummies		
	USD/EUR	JPY/USD	JPY/EUR	USD/EUR	JPY/USD	JPY/EUR
	Panel A. Fraction of volume with any computer participation, OLS estimation					
Coeff. on AT	0.0029	0.0018	0.0034^{***}	0.0078^{***}	−0.0030	0.0065^{***}
	(0.0024)	(0.0021)	(0.0012)	(0.0027)	(0.0024)	(0.0016)
R^2 (%)	53.44%	61.13%	71.90%	56.73%	62.57%	73.33%
	Panel B. Fraction of volume with any computer participation, IV estimation					
Coeff. on AT	-0.0121^{*}	-0.0186^{**}	−0.0022	−0.0078	−0.0101	−0.0128
	(0.0062)	(0.0089)	(0.0039)	(0.0061)	(0.0069)	(0.0175)
F-Stat	29.5800	19.4568	32.179	38.166	20.8947	2.2521
	Panel C. Fraction of volume with computer taking, OLS estimation					
Coeff. on AT	0.0037	−0.0027	0.0015	0.0094^{**}	−0.0034	0.0032^{**}
	(0.0036)	(0.0024)	(0.0012)	(0.0038)	(0.0027)	(0.0016)
R^2 (%)	53.39%	61.17%	71.56%	56.43%	62.55%	72.66%
	Panel D. Fraction of volume with computer taking, IV estimation					
Coeff. on AT	-0.0160^{**}	-0.0215^{**}	−0.0007	−0.0072	−0.0122	−0.0182
	(0.0080)	(0.0109)	(0.0028)	(0.0070)	(0.0082)	(0.0291)
F-Stat	39.9903	17.6348	64.8095	55.4489	21.2043	1.0441

Table 5: We report the mean ratio of the exchange rate volatility (Panel A), liquidity provision by humans and by computers (Panel B) estimated during announcement days divided by that estimated during non-announcement days. The one-hour measure is estimated using observations from 8:25 am to 9:24 am ET and the one-minute measure is estimated using 8:30 am to 8:31 am ET observations. Announcement days are defined as nonfarm payroll announcement days and non-announcement days are defined as 10 business days before and after the nonfarm payroll announcement. In each panel, we report the chi-squared and p-value of the Wald test that the ratio is equal to 1. In Panel C we report the chi-squared and p-value of the Wald test that the liquidity provision of humans during announcement days relative to non-announcement days is similar to the liquidity provision of computers. The statistics are estimated using data in the full sample from 2006 to 2007 and there are 23 observations (April 6, 2007 nonfarm payroll announcement is missing because it falls on Good Friday, when trading in the foreign exchange market is limited). Human liquidity provision, LH, is defined as the sum of human-maker/human-taker volume plus human-maker/human-taker volume divided by total volume. Computer liquidity provision, LC, is defined as the sum of computer-maker/computer-taker volume plus computer-maker/human-taker volume divided by total volume. The ***, **, and * represents significance at the 1, 5, and 10 percent level, respectively.

	USD/EUR		JPY/USD		JPY/EUR	
	Hour	Minute	Hour	Minute	Hour	Minute
	Panel A: Volatility					
$\frac{Vol_a}{Vol_n}$	6.236***	21.704***	5.595***	24.812***	3.697***	14.403**
χ^2 $(H_0 : Vol_a = Vol_n)$	69.86	18.76	33.34	15.45	19.37	5.96
p-value	0.0000	0.0003	0.0000	0.0008	0.0002	0.0235
	Panel B: Liquidity Provision					
Liquidity provision by humans, $\frac{LH_a}{LH_n}$	0.964***	1.062***	1.023	1.183***	0.888***	0.980
Liquidity provision by computers, $\frac{LC_a}{LC_n}$	1.132***	0.871***	0.974	0.652***	1.227***	1.151
χ^2 $(H_0 : LH_a = LH_n \text{ or } LC_a = LC_n)$	16.56	9.04	2.71	31.91	25.19	0.5
p-value	0.0005	0.0067	0.1143	0	0.0001	0.487
	Panel C: Comparison of Liquidity Provision between Humans and Computers					
$\frac{LH_a}{LH_n} - \frac{LC_a}{LC_n}$	−0.168***	0.191**	0.049	0.532***	−0.339***	−0.171
χ^2 $(H_0 : \frac{LH_a}{LH_n} = \frac{LC_a}{LC_n})$	19.24	5.91	1.50	36.07	25.21	0.66
p-value	0.0003	0.0241	0.2339	0.0000	0.0001	0.4245

Table 6: Impulse responses from the VAR specification with human-taker and computer-taker order flow. The table shows the impulse responses for returns as a result of shocks to the human-taker order flow ($HH+CH$) or computer-taker ($CC+HC$) order flow, denoted H-taker and C-taker in the table headings, respectively. The results are based on estimation of equation (2), using minute-by-minute data. In Panel A we show the return response, in basis points, to a one-billion base-currency shock to one of the order flows. In Panel B we show the return response, in basis points, to a one standard deviation shock to one of the order flows. We show the results for the full 2006-2007 sample and for the three-month sub-sample, which only uses data from September, October, and November of 2007. For each currency pair we show the short-run (immediate) response of returns; the corresponding cumulative long-run response of returns, calculated as the cumulative impact of the shock after 30 minutes; and the difference between the cumulative long-run response in returns minus the immediate response of returns, i.e., we provide the extent of over-reaction or under-reaction to an order flow shock. There are a total of $717,120$ minute-by-minute observations in the full two year sample and $89,280$ observations in the three-month sub-sample.

	Full 2006-2007 sample		3-month sub-sample	
	H-taker	C-taker	H-taker	C-taker
	Panel A: One billion base-currency shock			
	USD/EUR			
Short run	28.06	26.94	23.20	25.22
Long run	27.87	32.35	24.16	31.38
Difference	−0.20	5.42	0.96	6.16
	JPY/USD			
Short run	46.77	39.81	48.02	44.89
Long run	47.50	44.27	49.54	40.63
Difference	0.74	4.46	1.52	−4.26
	JPY/EUR			
Short run	99.32	102.71	124.02	115.52
Long run	108.07	109.85	132.53	123.26
Difference	8.75	7.14	8.51	7.74
	Panel B: One standard deviation shock			
	USD/EUR			
Short run	0.6617	0.2639	0.6045	0.3181
Long run	0.6570	0.3170	0.6296	0.3957
Difference	−0.0046	0.0531	0.0251	0.0777
	JPY/USD			
Short run	0.8706	0.3269	1.0241	0.5098
Long run	0.8843	0.3635	1.0565	0.4614
Difference	0.0137	0.0366	0.0324	−0.0483
	JPY/EUR			
Short run	0.5572	0.4901	0.7587	0.7636
Long run	0.6063	0.5242	0.8108	0.8148
Difference	0.0491	0.0341	0.0520	0.0512

Table 7: Variance decompositions from the VAR specification with human-taker and computer-taker order flow. The table provides the long-run variance decomposition of returns, expressed in percent and calculated at the 30 minute horizon, based on estimation of equation (2), using minute-by-minute data. That is, the table shows the proportion of the long-run variation in returns that can be attributed to shocks to the human-taker order flow ($HH + CH$) and the computer-taker ($CC + HC$) order flow, denoted H-taker and C-taker in the table headings, respectively. For each currency pair we show the actual variance decomposition, and the proportion of the *explained* variance in returns that can be attributed to each order flow type. That is, we re-scale the variance decompositions so that they add up to 100 percent. We show results for the full 2006-2007 sample and for the three-month sub-sample, which only uses data from September, October, and November of 2007. There are a total of $717,120$ minute-by-minute observations in the full two year sample and $89,280$ observations in the three-month sub-sample.

	Full 2006-2007 sample		3-month sub-sample	
	H-taker	C-taker	H-taker	C-taker
	USD/EUR			
Variance decomposition	29.27	4.74	25.92	7.25
Proportion of *explained* share	86.06	13.94	78.14	21.86
	JPY/USD			
Variance decomposition	29.31	4.22	28.59	7.22
Proportion of *explained* share	87.41	12.59	79.84	20.16
	JPY/EUR			
Variance decomposition	12.03	9.28	12.47	12.67
Proportion of *explained* share	56.45	43.55	49.60	50.40

Table 8: Impulse responses from the VAR specification with all four human/computer-maker/taker order flow combinations. The table shows the impulse responses for returns as a result of shocks to the human-maker/human-taker order flow (HH), computer-maker/human-taker order flow (CH), human-maker/computer-taker order flow (HC), or computer-maker/computer-taker order flow (CH), denoted in obvious notation in the table headings. The results are based on estimation of equation (3), using minute-by-minute data. In Panel A we show the return response, in basis points, to a one-billion base-currency shock to one of the order flows. In Panel B we show the return response, in basis points, to a one standard deviation shock to one of the order flows. We report the results for the full 2006-2007 sample and for the three-month sub-sample, which only uses data from September, October, and November of 2007. For each currency pair we show the short-run (immediate) response of returns; the corresponding cumulative long-run response of returns, calculated as the cumulative impact of the shock after 30 minutes; and the difference between the cumulative long-run response in returns minus the immediate response of returns, i.e., we provide the extent of over-reaction or under-reaction to an order flow shock. There are a total of 717,120 minute-by-minute observations in the full two year sample and 89,280 observations in the three-month sub-sample.

	Full 2006-2007 sample				3-month sub-sample			
	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker
	Panel A: One billion base-currency shock							
	USD/EUR							
Short run	27.64	29.66	26.57	32.19	20.58	30.94	28.94	21.74
Long run	30.13	20.47	29.89	24.92	24.18	23.35	34.64	5.94
Difference	2.49	−9.19	3.32	−7.26	3.60	−7.59	5.70	−15.80
	JPY/USD							
Short run	43.48	58.94	40.34	61.57	41.96	64.63	46.08	67.65
Long run	47.01	49.53	42.61	54.37	46.83	57.24	40.33	51.81
Difference	3.53	−9.41	2.27	−7.20	4.87	−7.39	−5.75	−15.85
	JPY/EUR							
Short run	102.61	92.16	100.91	102.04	139.33	103.92	114.01	94.47
Long run	116.12	91.24	107.18	93.41	159.46	96.85	118.47	95.20
Difference	13.51	−0.92	6.27	−8.63	20.13	−7.07	4.46	0.74
	Panel B: One standard deviation shock							
	USD/EUR							
Short run	0.5389	0.2575	0.2318	0.0893	0.4342	0.3211	0.3228	0.0845
Long run	0.5875	0.1777	0.2608	0.0692	0.5101	0.2424	0.3864	0.0231
Difference	0.0486	−0.0798	0.0290	−0.0202	0.0760	−0.0788	0.0636	−0.0614
	JPY/USD							
Short run	0.6721	0.3968	0.2962	0.1506	0.7019	0.5801	0.4544	0.2607
Long run	0.7267	0.3334	0.3129	0.1330	0.7834	0.5137	0.3976	0.1997
Difference	0.0546	−0.0634	0.0167	−0.0176	0.0815	−0.0663	−0.0567	−0.0611
	JPY/EUR							
Short run	0.4440	0.2629	0.4481	0.1583	0.5859	0.3829	0.6809	0.2409
Long run	0.5024	0.2603	0.4760	0.1449	0.6706	0.3568	0.7076	0.2428
Difference	0.0584	−0.0026	0.0279	−0.0134	0.0847	−0.0260	0.0266	0.0019

Table 9: Variance decompositions from the VAR specification with all four human/computer-maker/taker order flow combinations. The table provides the long-run variance decomposition of returns, expressed in percent and calculated at the 30 minute horizon, based on estimation of equation (3), using minute-by-minute data. That is, the table shows the proportion of the long-run variation in returns that can be attributed to shocks to the human-maker/human-taker order flow (HH), computer-maker/human-taker order flow (CH), human-maker/computer-taker order flow (HC), and computer-maker/computer-taker order flow (CH), denoted in obvious notation in the table headings. We show the actual variance decomposition, and the proportions of the *explained* variance in returns that can be attributed to each order flow type. That is, we re-scale the variance decompositions so that they add up to 100 percent. We present results for the full 2006-2007 sample and for the three-month sub-sample, which only uses data from September, October, and November of 2007. There are a total of $717,120$ minute-by-minute observations in the full two year sample and $89,280$ observations in the three-month sub-sample.

	Full 2006-2007 sample				3-month sub-sample			
	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker
	USD/EUR							
Variance decomp.	20.71	4.73	3.89	0.58	14.19	7.68	7.86	0.59
Proportion	69.24	15.81	13.01	1.94	46.80	25.33	25.92	1.95
	JPY/USD							
Variance decomp.	18.62	6.48	3.70	0.93	14.47	9.78	6.12	2.00
Proportion	62.63	21.80	12.45	3.13	44.70	30.21	18.91	6.18
	JPY/EUR							
Variance decomp.	7.84	2.74	7.94	0.99	7.72	3.32	10.47	1.30
Proportion	40.18	14.04	40.70	5.07	33.84	14.56	45.90	5.70

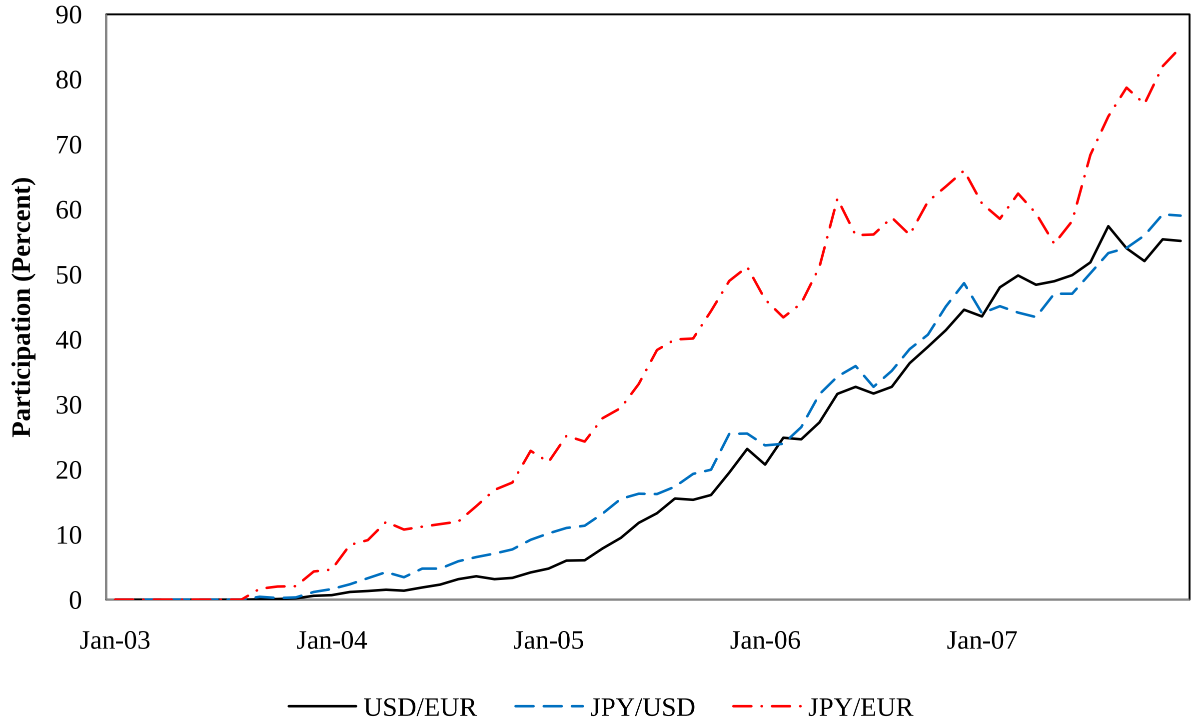

Figure 1: Participation rates of algorithmic traders

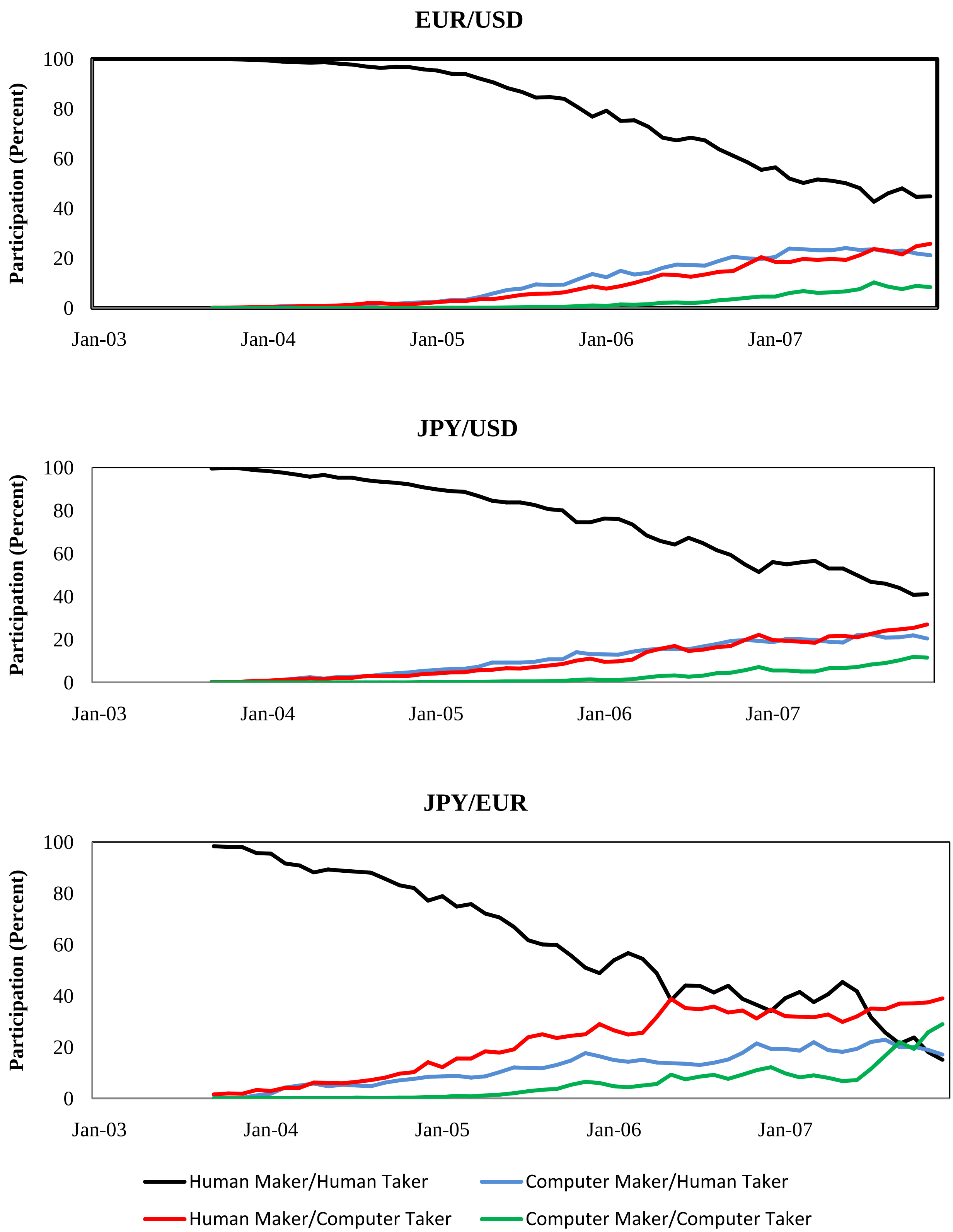

Figure 2: Participation rates broken down into four maker-taker pairs

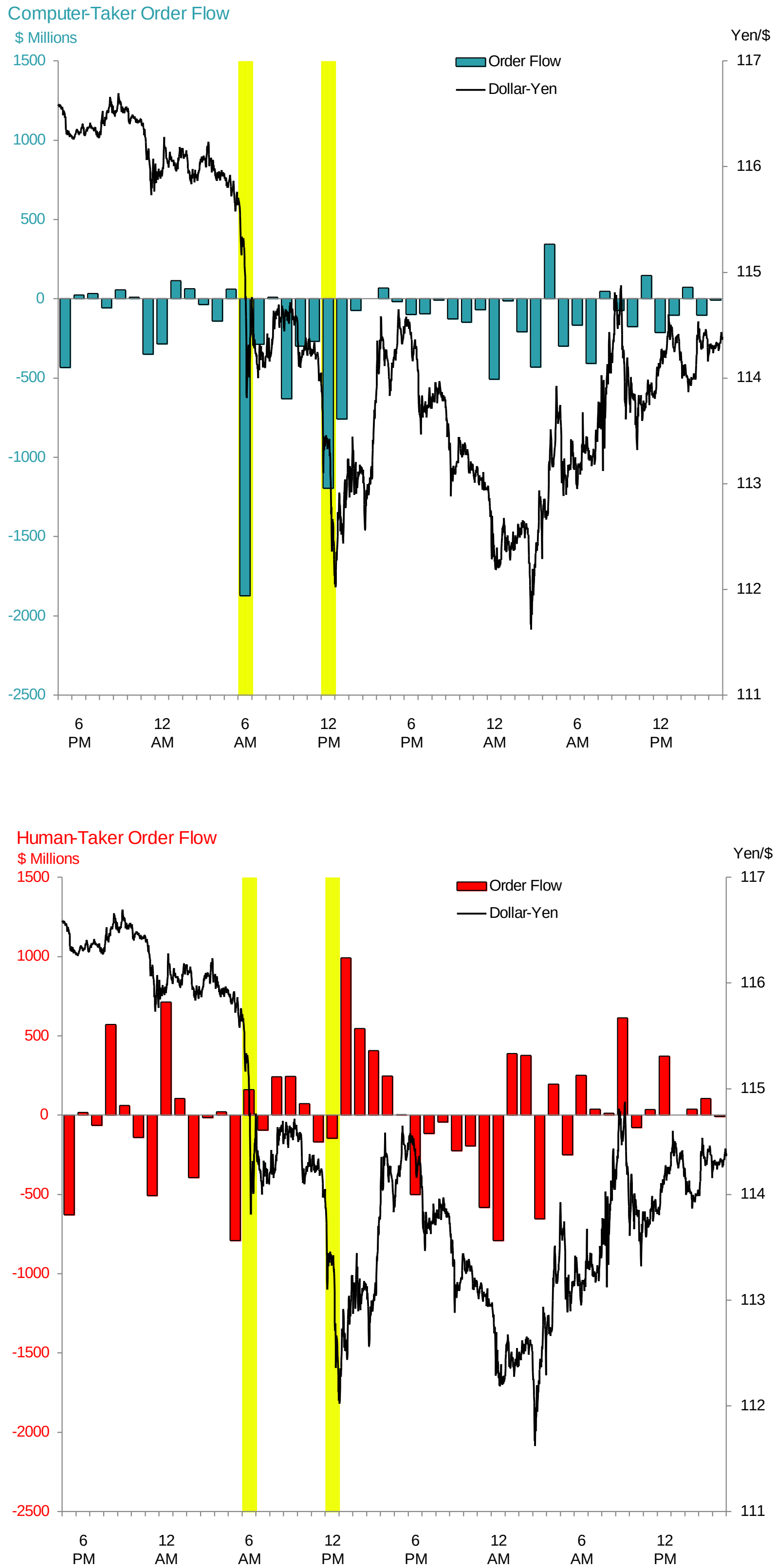

Figure 3: Dollar-Yen Market on August 16, 2007

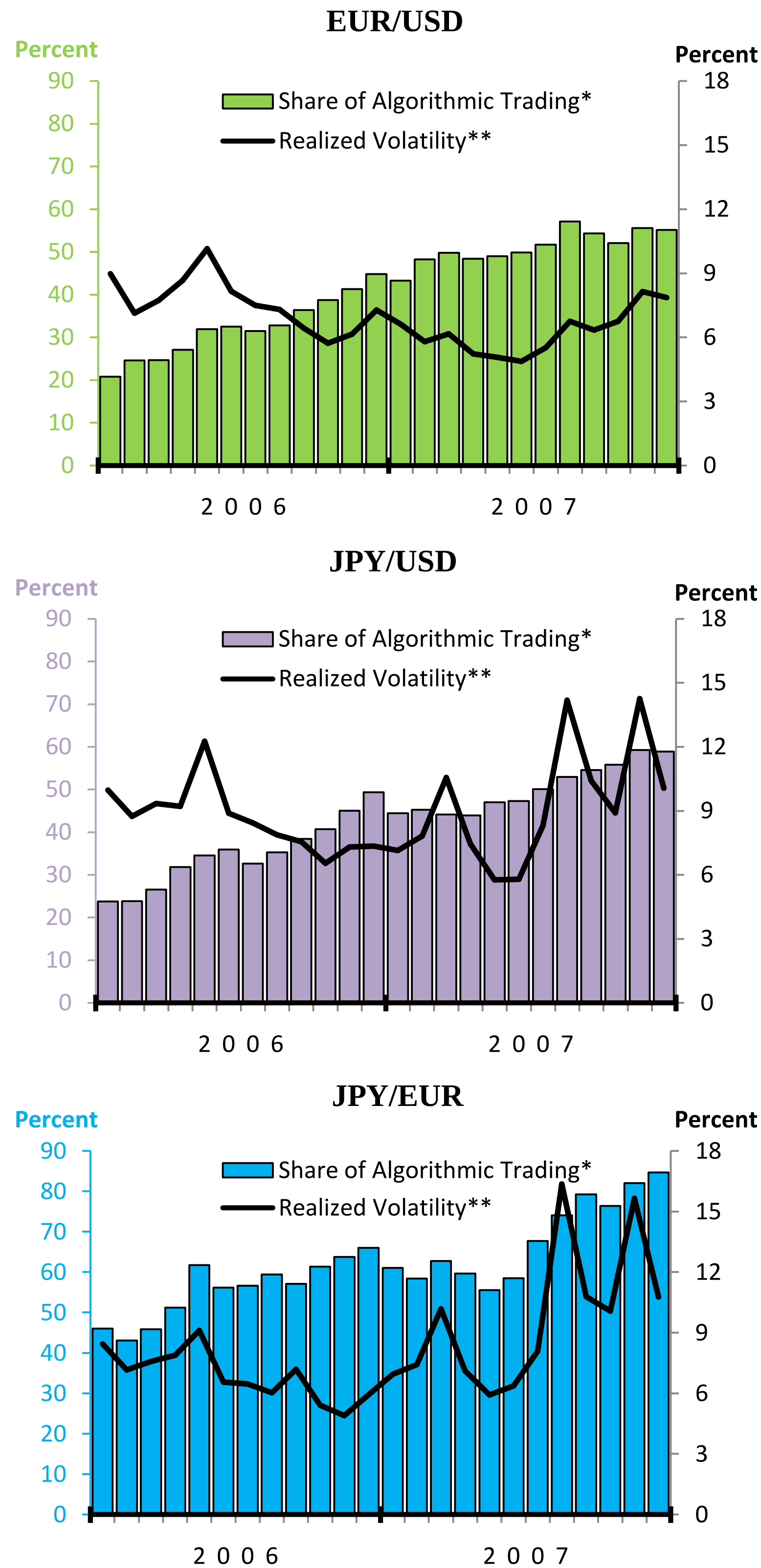

Figure 4: Volatility and Algorithmic Market Participation

*Daily realized volatility is based on 5-minute returns. We show monthly observations
**Share of algorithmic trading is at a monthly frequency