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

Alain Chaboud Benjamin Chiquoine Erik Hjalmarsson Clara Vega*

[PRELIMINARY AND INCOMPLETE]

December 17, 2008

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 five major currency pairs from 2006 to 2007. Our dataset contains precise observations of the fraction and the direction of the computer-generated trades each minute. As such, it allows us to analyze the possible links between algorithmic trading and market volatility, to identify whose trades have a more permanent impact on prices, and to study how correlated algorithmic trades are. 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". We also find that there is, in some cases, an over-reaction of the price to algorithmic order flow. Thee is some evidence that algorithmic trades tend to be correlated, suggesting that the algorithmic strategies used in the market may not be as diverse as those used by non-algorithmic traders.

JEL Classification: F3, G12, G14, G15.

Keywords: Algorithmic trading, excess volatility, private information.

*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. 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.

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, however, there has been almost no 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]

In the foreign exchange market, the adoption of algorithmic trading (AT) 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 AT has been rapid, however, and a sizable fraction of foreign exchange transactions currently involve at least one algorithmic counterparty. We study in this paper the impact that AT has had on price discovery and volatility using high-frequency data representing a majority of global interdealer trading in five major currency pairs from January 2006 to December 2007, a period over which the share of AT in the foreign exchange markets rose rapidly. Importantly, our dataset contains precise observations of the fraction and the direction of the computer-generated trades each minute. As such, it allows us to study how algorithmic trading and market volatility are related, to identify whether algorithmic or non-algorithmic trades have a more permanent impact on prices, and to estimate how correlated algorithmic trades are.

In algorithmic trading, 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, 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

[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.

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, as some have feared that AT may generate large price swings and market instability. One such instance may have happened on August 16, 2007, in a period of extreme volatility, the highest in 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 we show in Figure 1. The figure also shows, for each 30-minute interval in the day, algorithmic ("computer") order flow in the top panel and non-algorithmic ("human") order flow in the lower panel. The two sharp exchange rate movements mentioned happened when computers, as a group, aggressively sold dollars and bought yen. Human order flow at those times was, in contrast, small. Humans traders then aggressively bought dollars after 12:00 p.m, 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 whether computer trades tend to create excess volatility, in the sense that exchange rate movements driven by computer trades are more likely to be later reversed. This example also leads us to ask whether human trades routinely have a more permanent impact on prices than computer trades.

We formally investigate these conjectures using minute-by-minute data from January 2006 to December 2007 on five exchange rate pairs: the euro-dollar, dollar-yen, dollar-swiss franc, euro-yen, and euro-swiss franc. We find that, controlling for potential endogeneity biases and for the common trend in exchange rate volatility and algorithmic trading, there is no evident causal relationship between AT and volatility. However, the instruments we use in the analysis are weak and thus we also analyze return-order flow dynamics in a high-frequency VAR framework in the tradition of Hasbrouck (1991a).

The VAR estimation provides three important insights. First, we find that human order flow accounts for most of the (long-run) variance in exchange rate returns, i.e., humans are the “informed” traders in these markets. 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.

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 this is evidence that computers tend to place limit orders more strategically than humans do. This finding may relate to the literature that proposes to depart from the

[3] The Economist, June 21, 2007

prevalent assumption that liquidity providers in limit order books are passive.[4]

Third, our VAR analysis shows that there is an initial under-reaction to order flow between humans (where the price is both posted by and dealt on by a human), while there is an initial over-reaction to order flow between computers. The euro-dollar exchange-rate pair during our three-month subsample provides an extreme example, as the initial reaction to computer-computer order flow is a 21 basis point move, but the long-run cumulative reaction is just 4 basis points. To the extent that there is an initial over-reaction to computer-computer order flow, we conclude that algorithmic trading may be linked to some excess short-run volatility. Also, in the euro-dollar and dollar-yen markets, the presence of a computer as a liquidity provider or a liquidity demander is linked to some short-term overreaction of the price. But this is not the case for the dollar-swiss, the euro-yen, and the euro-swiss franc exchange rates. Coincidentally, these are the exchange rates where, by the end of our sample, AT is as prevalent or more prevalent than human trading. We believe that a substantial fraction of the AT in these markets reflects computers taking advantage of so-called triangular arbitrage opportunities, where the prices set in, say, the euro-dollar and dollar-yen markets are very briefly out of line with the eur-yen rates. In these cases, computer trading likely contributes to market efficiency by narrowing misalignments of exchange rates.

Finally, we find some evidence that, in all our currency pairs, 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. This fact echoes the concerns voiced by some analysts that, as computers take over trading in financial markets, these markets will miss the benefits of the divergence of opinion among humans, as well as their slower reaction times and perhaps more subtle judgment. However, since the high correlation of computer trades does not seem to automatically translate into economically significant excess volatility, it is not clear how damaging that high correlation is.

We proceed as follows. In Section 2 we describe the Electronic Broking Services (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 relationship between realized volatility and the activity of algorithmic traders. The econometric techniques used in this section take advantage of differences in the time series of volatility and AT prevalence among the different currency pairs to address likely endogeneity issues. In Section 4 we analyze return-order flow dynamics in a VAR framework to identify whose trades, computers or humans, have a more permanent impact on prices. In Section 5 we examine how correlated the algorithmic orders are with each other compared to the human orders. In Section 6 we conclude.

[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.

2 Data description

Today, two electronic platforms process the vast majority of global interdealer spot trading in all the major currency pairs, one offered by Reuters, and one offered by EBS. 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 become over time 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, for example, 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.

We have access to AT data from EBS from 2003 through 2007. We focus, however, on the sample from 2006 to 2007, because, as we show in Figure 2, 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. Since the growth in algorithmic trading continued throughout 2006 and 2007, it is interesting to separately analyze these three months towards the end of the sample period, when algorithmic trading played an even more important role than earlier in the sample.[5] We have access to the five most-traded currency pairs on the EBS system: euro-dollar, dollar-yen, euro-yen, dollar-swiss franc, and euro-swiss franc. The quote data, at the one-second frequency, consist of the highest bid quote and the lowest ask quote on the EBS system in these currency pairs, from which we construct mid-quote series and compute one-minute exchange rate returns. The transactions data, at the one-minute frequency, consist, for each currency pair, of the amounts of base currency bought and sold. We can also identify the type of trader, human or computer, who posted the price at which the transaction was conducted (the "maker") and the type of trader who decided to buy or sell at that price (the "taker").[6]

The main goal of this paper is to analyze the effect algorithmic trading has on price discovery and volatility. To that end, we analyze different decompositions of "order flow" (we clearly stretch the traditional definition of order flow, as shown below). First, we decompose order flow into the four most disaggregate components: human-maker/human-taker (HH), computer-maker/human-taker (CH), human-maker/computer-taker (HC), and computer-maker/computer-taker (CC). Second, we decompose order flow into the standard separation, which distinguishes trades based on who initiated the trade: human-taker (HH+CH) and computer-taker (HC+CC). Third, we decompose order flow into a separation that distinguishes trades based on who provides liquidity: human-maker (HH+HC) and computer-maker (CH+CC). Fourth, we decompose order flow into

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

[6] 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 time. Order flow measures based on amounts transacted and those based on number of trades are therefore very similar.

purely human trades (HH) and trades where at least one of the two counterparties was an algorithmic trader (CH+HC+CC).

The first decomposition allows us to analyze 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-rates, where computers likely have a clear advantage over humans in detecting short-lived triangular arbitrage opportunities. This decomposition may also allow us to study whether the liquidity supplier, who is traditionally assumed to be “uninformed”, is 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.[7] Still, in our exchange rate data as in other financial data, the net of trades signed by who the taker is (the standard definition of order flow) is clearly highly positively correlated with exchange rate returns, so that the taker is considered to be more "informed" than the maker. Thus we also consider prominently the more traditional second decomposition, human-taker and computer-taker order flow, in our analysis of whose trades impact prices more. The third decomposition, human-maker and computer-maker order flow, allows us, for instance, to determine whether computers or humans are more likely to provide liquidity when it is needed the most, e.g. during periods of high exchange rate volatility. Lastly, the fourth decomposition, computer-participation and human-human order flow, allows us to determine whether any type of participation by computers, passive or active, is linked to excess volatility in the market.

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 Independence Day holiday is observed).

We show summary statistics for the one-minute returns and order flow data in Table 1 and Table 2. These tables contain a number of noteworthy features. First, order flow is serially positively correlated. This is consistent with informed trading models. For example, Easley and O’Hara (1987) model a situation where sequences of large purchases (sells) 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

[7] 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.

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 noninformational reasons and splitting his order to reduce market impact. The serial correlation in order flow is of additional interest to us because we will need to control for it in our regression specifications. Second, the standard deviation of order flow is different across maker/taker pairs and exchange rates. For example, in the two exchange rate markets with the highest trading volume, the euro-dollar and dollar-yen markets, the standard deviation of human-taker order flow is larger than the standard deviation of computer-taker order flow. A consequence is that, in these two markets, large one-sided market orders are more likely to be executed by human takers than by computer takers. Differences in standard deviations across maker/taker order flow pairs are important to the interpretation of the VAR analysis. It could be the case, for instance, that the price impact of a one billion dollar shock to CC order flow is the same as the price impact of a one billion dollar shock to HH order flow. However the percent of the total variation in exchange rate prices explained by the latter type of order flow would be larger because its standard deviation is larger.

The correlations between the most disaggregate types of order flow are shown in Table 3, both for the full 2006-2007 sample as well as for the shorter three month sub-sample. One notable result in these tables is that the four types of order flow are not highly correlated (positively or negatively) except for the HC and CH order flows, with a correlation of about -0.4. This is consistent with Parlour and Seppi (2008)'s assertion that, in a limit order book, investors with active trading motives may choose to place limit orders that are more agressive than those a disinterested liquidity provider would place. In other words, when computers, for example, want to buy dollars and sell euros they will not only do it by executing market orders but they will also post limit orders that are aggressive and are more likely to be picked up by humans than by other computers, i.e. when HC is positive CH tends to be negative.[8]

We show in Figure 2, from 2003 through 2007, for the five major currency pairs trading on EBS, the fraction of trades where at least one of the two counterparties was an algorithmic trader (CH+HC+CC). From its beginning in the second half of 2003, the fraction of trades involving AT grew by the end of 2007 to near 60% for euro-dollar, dollar-yen, and euro-swiss trading, and to about 80% for euro-yen and dollar-swiss. Figure 3 shows, for each of our five currency pairs, the evolution over time of the four different possible types of trades. By the end of 2007, in the euro-dollar market, 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 percent. Computers "took" prices posted by humans about as often as humans took prices posted by market-making computers, in blue. The same pattern is also found in the dollar-yen market. Since the presence of more

[8]We also note that, since the correlation across different types of order flow is not extremely high, with perhaps the exception of CH and HC, we can have less concern about multicollinearity in some of our regression specifications.

"makers" increases market liquidity, i.e., larger trades can be executed with little impact on the price, Figure 3 shows that in the most-traded currency pairs, computer and human traders contributed about evenly to market liquidity. 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, and the most common type of trade was computers trading on prices posted by humans (HC). We believe this reflects computers taking advantage of 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. Trading volume is largest in euro-dollar and dollar-yen markets, and price discovery happens mostly in those markets, not in the cross-rate. The dollar-swiss franc and euro-swiss franc markets are also more highly reliant on AT by the end of 2007.[9]

3 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.

3.1 Identification

The main challenge in identifying a causal relationship between algorithmic trading and volatility is the potential endogeneity of algorithmic trading with regards to variables such as volatility. That is, although one may conjecture that algorithmic trading impacts volatility, it is also highly plausible that algorithmic trading activity is 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. The bottom line is that the fraction of algorithmic trading is likely to be endogenous with regards to exchange rate volatility. We cannot easily determine in what direction the bias will go in an OLS regression of volatility on the fraction of algorithmic trading, because it is not obvious whether higher volatility would induce more or less algorithmic trading. To deal with the endogeneity issue, we adopt an instrumental variable (IV) approach as outlined below.

[9] The foreign exchange markets for the Swiss franc are highly dependent on the trading activity of the two large Swiss banks, UBS and Credit Suisse, which are known for their sophisticated electronic trading activity. Traders tell us that, in the Swiss franc exchange markets, the dollar-swiss franc pair is generally viewed as the "third leg" of the triangular arbitrage play, with price discovery occuring primarily in euro-dollar and euro-Swiss franc.

We are interested in estimating the following regression equation,

$$RV_{i,t} = \alpha_i + \beta_i AT_{i,t} + \gamma_i' x_{i,t} + \epsilon_{i,t}, \tag{1}$$

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 the fraction of algorithmic trading at time t in currency pair i, $x_{i,t}$ is a set of control variables that will primarily contain lagged values of $RV_{i,t}$ as well as 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 $x_{i,t}$ but not necessarily with $AT_{i,t}$. The exact definitions of $RV_{i,t}$, $AT_{i,t}$, and $x_{i,t}$ will be given later.

The main focus of interest is the parameter β_i, which measures the impact of algorithmic trading on $RV_{i,t}$ 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). Thus, $z_{i,t}$ needs to be uncorrelated with the variation left in volatility, after controlling for the variation in $x_{i,t}$ and $AT_{i,t}$.

The starting point of our identification scheme is the fact that we have data on several currency pairs. A natural instrument for $AT_{i,t}$ that comes to mind is therefore algorithmic trading in the other currency pairs $\{AT_{j,t}\}_{j \neq i}$. However, since volatility is correlated across currency pairs, it is likely that $\epsilon_{i,t}$ is also correlated across currency pairs. Thus, under the assumption that $AT_{i,t}$ and $\epsilon_{i,t}$ are correlated, it follows that it is likely that $AT_{j,t}$ and $\epsilon_{i,t}$ are also correlated and $AT_{j,t}$ may therefore not be a valid instrument.

Instead, we propose to use the lagged values of $AT_{j,t}$ as instruments; that is, $\{AT_{j,t-1}\}_{j \neq i}$. Since there is both serial correlation and cross-correlation across currencies in the fraction of algorithmic trading, these instruments should be relevant; i.e., correlated with $AT_{i,t}$. Importantly, however, these lagged variables are also likely to be valid instruments when $x_{i,t}$ is defined appropriately. For instance, let $x_{i,t}$ include lagged values of both $RV_{i,t}$ and $\{RV_{j,t}\}_{j \neq i}$. The lags of own volatility are used to control for the well known serial correlation in volatility. The lags of the volatility for the other currency pairs are included to ensure the validity of the proposed instruments. That is, by controlling for lagged values of volatility in *all* currency pairs, the error term $\epsilon_{i,t}$ should only be *contemporaneously* correlated with the volatility in other currencies, and not with the lagged values. Consequently, we would also expect $\epsilon_{i,t}$ to be uncorrelated with the lagged values of algorithmic trading in other currencies, $\{AT_{j,t-1}\}_{j \neq i}$, which suggests that these should provide valid instruments. Since the cross-currency exchange rates (the euro-franc and the euro-yen) are effectively determined by the other three main currency pairs in the sample, there might be some concern that this would affect the validity of the above IV approach. In addition to performing the IV estimation using all five currency pairs, we therefore also repeat the estimation using only the three main currency pairs.

The instrumental variable regressions are estimated using Limited Information Maximum Likelihood (LIML), and we test for both the relevance and the validity of the instruments by reporting the Stock and Yogo (2005) test of weak instruments and the standard $J-$test of overidentifying restrictions, which provides a test of the instrument validity. We use LIML rather than two-stage OLS since Stock and Yogo (2005) show that the former is much less sensitive to weak instruments than the latter.

Another inferential issue, quite distinct from the endogeneity issue just discussed, is the strong upwards trend in the fraction of algorithmic trading over the sample period. As seen in the previous graphs, this trend is clearly the dominant feature of the time-series behavior of the fraction of algorithmic trading. Thus, if one does not attempt to control for it, any regression results with algorithmic trading as an explanatory variable will primarily reflect the correlation between the left-hand-side variable and this increasing secular trend. For instance, if there is a tendency for the left-hand-side variable to trend downwards over the sample period, as is the case for the volatility in some currency pairs, then the estimated slope-coefficient will most likely be negative. However, although one cannot rule out that there is therefore a long-run negative relationship between volatility and algorithmic trading, it is also quite possible that the downward trend in volatility is driven by some other factor that is not accounted for in the model. Since there is no feasible way to control for all other potential factors that may have caused long-term shifts in the level of volatility, we focus on the impact of changes in algorithmic trading from some local mean that changes over time. In particular, monthly time dummies are included as control variables. The regression results should therefore be interpreted as the impact of changes in algorithmic trading over shorter time periods and not the effect of going from virtually no algorithmic trading to, say, $30-40$ percent of the total trading volume. The latter question is arguably as interesting as the former, but extremely difficult to answer without very strong assumptions.

3.2 Variable definitions

3.2.1 Realized Volatility

Volatility is measured as the *daily realized volatility* obtained from five minute returns; that is, the volatility measure is equal to the daily sum of squared five 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. We use five minute returns to avoid any bias in the estimation of volatility, which may arise from market microstructure noise present in returns sampled at even higher frequencies (e.g. Hansen and Lunde, 2006). Following the common conventions in the literature on volatility modelling (e.g. Andersen et al., 2001), the realized volatility is log-transformed to obtain a more well behaved time-series; naturally, all lags of volatility used in the regressions are also log-transformed.

3.2.2 Algorithmic trading

The amount of algorithmic trading is measured as the percent of the overall trading volume that includes an algorithmic trader as either a maker or a taker; that is, the percent of trading volume where a computer was involved in at least one side of the trade. In addition, we also considered an alternative measure that separates the trading volume into our four different types of trades, and calculates the percent of total volume that each type represents. The four different types, as before, are the trades where both maker and taker are human, where the maker is a human and the taker is a computer, where the maker is a computer and the taker is a human, and where both maker and taker are computers. However, using this finer measure of algorithmic trading added little to the empirical results found for the simpler measure and we do not report those results.

3.2.3 Other control variables

The additional control variables included in the regressions, represented by $x_{i,t}$ in equation (1), are discussed below. First, lagged values of the dependent variables are included to control for serial correlation. Realized volatility has a strong serial correlation even for distant lags (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003 and Bollerslev and Wright, 2000). We follow the work of Andersen et al. (2007). In particular, to control in a parsimonious manner for the serial correlation in volatility, which tend to stretch back many lags, we include the first daily lag of volatility, the weekly lag of volatility, calculated simply as the average over the past five business days and the monthly lag of volatility, calculated as the average over the past 22 business days. As argued by Andersen et al. (2007), such a lag structure will capture most of the 'long-memory' features of (logged) realized volatility, without imposing a vast number of parameters to estimate.[10]

Second, as described in the context of the instrumental variable approach outlined above, the lagged values of realized volatility for the other currency pairs are also included as regressors; the same number of lags as for the own dependent variable are used in all regressions. Finally, in order to control for the large secular trends in the fraction of algorithmic trading, monthly time dummies are included in the regressions.

[10] Alternatively, one could also model the realized volatility as a long-memory or fractionally integrated process. In this case, the long-memory parameter (d) is estimated and the fractionally differenced realized volatility series is used in the analysis. The results from such a specification are qualitatively identical to those shown in the paper, and are not presented.

3.3 Empirical results

The empirical regressions results are presented in Table 4. We present OLS results, the LIML-IV results using all five currencies, and the LIML-IV results using only the three main currency pairs. Each specification is estimated with or without time dummies as outlined previously. The lag structure described above is included in all regressions. We report results for the sample starting in January 2006 and ending in December 2007. In order to save space, only the estimates of the coefficient in front of the fraction of algorithmic trading are presented. As described before, the specification with time dummies represents the most interesting and relevant one, whereas the one without time dummies is primarily included for completeness.

In addition to the coefficient on the fraction of algorithmic trading, the results for the $J-$test of overidentifying restrictions, which provides a test of the instrument validity, and the Stock and Yogo (2005) $F-$test of weak instruments, which tests instrument relevance, are reported for the IV regressions. Failure to reject with the $J-$test provides some evidence of the validity of the instruments. The Stock and Yogo (2005) $F-$statistic, which is equivalent to the $F-$statistic for the excluded instruments in the first stage regression, tests whether the instruments are weak. 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 with all five currencies used in the IV estimation, a value greater than 5.44 for this $F-$statistic indicates that the maximal size of a 5 percent test will be no greater than 10 percent, which might be deemed acceptable; the corresponding critical value in the three currency specification is 8.68. In general, the larger the $F-$statistic, the stronger the instruments.

The OLS results show that there appears to be a positive *association*, or correlation, between the level of volatility and the fraction of algorithmic trading in the market, with highly significant estimates in all but one currency. The OLS estimates are not likely to provide an unbiased estimate of the causal relationship and turning to the IV results, most signs of a relationship disappears. The $F-$statistics for the IV estimation raise some warning signs, however. In the specification with time dummies, which is of primary interest, the null of weak instruments can typically not be rejected at a level that insures no more than a maximal size of 10 percent in the tests of the IV coefficient. Only for the euro-dollar currency pair are there no substantial signs of weak instruments and in this case the coefficient on algorithmic trading is insignificant; the coefficient on algorithmic trading is insignificant for all other currencies as well, when time dummies are included.

4 The price impact of algorithmic trading

In the previous section, we investigated whether the presence of algorithmic trading increases exchange rate volatility. However the inference is complicated by the secular trend in both algorithmic trading and realized volatility, as well as by endogeneity complications. In this section we indirectly determine whether computer trades cause excess volatility and high-frequency noise in the exchange rate. To this end we estimate return-order flow dynamics in a vector autoregressive (VAR) framework in the tradition of Hasbrouck (1991a). This procedure allows us to identify whose trades, computer or human, have a permanent impact on prices and to determine whether exchange rate prices are more likely to over-react to computer or human trades. We will interpret the price's over-reaction to a particular type of order flow as this particular type of order flow causing excess volatility. This interpretation is consistent with information-based models (dynamic learning models with informed and uniformed investors), where liquidity traders do not contribute to the price discovery process (do not have a permanent impact on prices) and prices temporarily over-react to this type of trades (e.g., Albuquerque and Miao (2008)), thus create excess volatility in prices. In contrast, in these models, the under-reaction of asset price's to order flow is a natural consequence of the learning process.[11]

4.1 VAR estimation

Similar to Hasbrouck (1991a), we allow returns to be contemporaneously affected by order flow, but there is no contemporaneous effect of returns on order flow. We also allow U.S. macroeconomic news surprises to affect both returns and order flow (Evans and Lyons, 2008). In particular, we estimate the following system of equations for each currency i

$$\begin{aligned} r_{it} &= \alpha^r + \sum_{j=1}^{J} \beta_{ij}^r r_{it-j} + \sum_{l=1}^{L}\sum_{j=0}^{J} \gamma_{ijl}^r OF_{it-j}^{(l)} + \sum_{k=1}^{K} \delta_{ik}^r S_{kt} + \varepsilon_{it}^r, \qquad (2) \\ OF_{it}^{(l)} &= \alpha_l^{OF} + \sum_{j=1}^{J} \beta_{ijl}^{OF} r_{it-j} + \sum_{l=1}^{L}\sum_{j=1}^{J} \gamma_{ijl}^{OF} OF_{it-j}^{(l)} + \sum_{k=1}^{K} \delta_{ikl}^{OF} S_{kt} + \varepsilon_{it}^{OF^{(l)}}. \end{aligned}$$

where $L = 2$ or 4 depending on the decomposition of the order flow; that is, $OF_{it}^{(l)}$ represents the l'th component of order flow in currency i at time t, where the order flow components are specified in the decompositions below. r_{it} is the 1-minute exchange rate return for currency i at time t; OF_{it} is the currency i order flow

[11] 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. In particular, order flow types are not public knowledge, so that agents cannot condition on these variables.

at time t decomposed in three different ways: defined in the most disaggregate way, $\{HH, HC, CC, CH\}$, defined according to who initiates the trade, $\{HH + CH, CC + HC\}$, and defined according to whether there is any computer participation in the market, $\{HH, CC + HC + CH\}$; S_{kt} is the macroeconomic news announcement surprise for announcement k defined as the difference between the announcement realization and its corresponding market expectation. We use the International Money Market Services (MMS) Inc. 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 listed in Table 5 and are similar to those in Andersen et al. (2003, 2007) and Pasquariello and Vega (2007).[12] For a detailed description of the data we refer the reader to Andersen et al. (2003). 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 ommitted variable bias in our specification to be small.

Following the tradition of the VAR price-impact literature we focus on the highest sample frequencies and estimate the VARs using the minute-by-minute data. 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 decomposition, it is worth briefly summarizing the main lessons from the estimated coefficients in the VAR. Since there are many coefficients estimated for each currency pair, we only report in Table 5 the macroeconomic news announcement coefficients and the contemporaneous order flow coefficients in the exchange-rate equation when we consider the most disaggregate decomposition of order flow: HH, CH, HC, and CC. The rest of the coefficient estimates are not shown but we briefly summarize our results below and we also report the impulse response function results. In addition to the coefficient estimates we report the R^2 of estimating the structural VAR with OLS equation by equation and the R^2 when we only consider news announcement times and run an OLS regression of 1-minute exchange rate returns on macroeconomic news announcements. This latter R-squared indicates how much do U.S. macroeconomic news announcements affect exchange rate returns. As theory would predict, we find that U.S. macroeconomic news announcements affect less the euro-swiss and the euro-yen than the euro-dollar, dollar-yen and dollar-swiss franc exchange rates.

[12] 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.

The first own lag in all the order flow equations is always highly significant, and typically around 0.1 for all currency pairs. The main exception is the coefficient on the own lag in the computer-maker/computer-taker order flow regression, where the first order autoregressive coefficient is typically much smaller and in the range 0.01 to 0.05. There is thus a sizeable first-order autocorrelation in most of the order flow components, but less so in the computer-maker/computer-taker order flow; the higher order lags are generally substantially smaller, but typically positive. The coefficients on the first order cross-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 human-maker/human-taker order flow, suggestive of a form of 'trend chasing' in the order placement. Interestingly, the opposite is true for the computer-maker/computer-taker order flow, where the first lag of returns always has a negative coefficient; for the other two order flows, the results are mixed across currencies.

Finally, the return equation shows that minute-by-minute returns tend to be negatively serially correlated, with the coefficient on the first own lag varying between -0.05 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. All four order flows are significant predictors of returns. The price impact of the lagged order flows range from around 1 to 15 basis points per billion units of order flow (denominated in the base currency), as compared to a range of approximately $20 - 100$ basis points in the contemporaneous order flow. The main differences in the coefficients on the lagged order flows in the returns equation are between currencies rather than between the different types of order flows. That is, there is little evidence that one type of order flow is a better predictor of returns than the others. Again, the first order lags dominate the relationship.

It should also be stressed that despite the strongly significant estimates that are recorded in the VAR estimations and the relatively high R^2 reported in Table 5, 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 four percent for the order flow equations, and often less than one percent for the return equations. Again, the main exception is the computer-maker/computer-taker order flow equation, which typically yields R^2s of less than one percent.

Overall, from examining the coefficients in the estimated VARs, there is little evidence that there is any systematic difference between the different types of order flows in the way that they affect the dynamics of returns. The most notable finding is probably the substantially lower persistence and predictability that is found for the computer-maker/computer-taker order flow.

4.2 Impulse Response Function Results

In Table 6 we show a summary of the results from the impulse response analysis based on the full sample for 2006-2007, when the size of the shock is the same across the different types of order flow: one billion base currency shock to order flow. Because the standard deviation of order flow is different across maker/taker order flow pairs (shown in Tables 1 and 2) we also consider a shock that varies across the different maker/taker order flow pairs according to the average size shock. To that end we show in Table 7 the (cumulative) impulse response of returns to one standard deviation shock to a particular type of order flow. All the responses are measured in basis points. Each table has three panels in which we show the results from the three different order flow decompositions: defined in the most disaggregate way, $\{HH, HC, CC, CH\}$, defined according to who initiates the trade, $\{HH + CH, CC + HC\}$, and defined according to whether there is any computer participation in the market, $\{HH, CC + HC + CH\}$. We show the short-run (instantaneous) impulse responses as well as the long-run cumulative responses, along with the long-run variance decomposition (Table 10 and 11). The long-run statistics are all 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.

Figures 4 and 5 trace out the full paths of the cumulative impulse responses based on the most disaggregate decomposition of order flow, again using two different shock sizes: one billion base currency order flow shock and one standard deviation shock, respectively. Given the large sample sizes being used, there is little gain from showing confidence bands for the impulse response functions, as the coefficients are very precisely estimated. In general, the standard errors of the estimates are small enough that they contribute little to the analysis, and are not displayed in the tables either.

Starting with a hypothetical shock of one billion base currency order flow, the results in Table 6 and Figure 4 show that, in general, HH order flow impacts prices more than CC order flow. However the differences in price impact, although statistically significant, are not economically significant. In particular, the difference in the responses across order flow types in the two currencies with the largest trading volume, the euro-dollar and dollar-yen markets, is very small, it ranges from 10 to 0 basis points. Furthermore, there are some notable exceptions to this pattern. For instance the immediate response of the dollar-yen exchange rate to a CC order flow shock is almost 50 percent larger than the response to a HH shock of the same size. For the euro-swiss franc, the opposite is true. A similar picture emerges when we decompose order flow according to who initiated the trade, human-taker compared to computer-taker, and according to whether there is any computer participation, human-maker/human taker compared to computer-participation.

In contrast to these results, the response to a hypothetical one standard deviation shock to the different order flows consistently shows that humans have a bigger impact on prices than computers (Table 7 and

Figure 5) and the differences are economically significant. In particular, one standard deviation shock to HH order flow has an average long-run effect of 0.5 basis points across currencies compared to one standard deviation shock to CC order flow which has an average effect of 0.1 basis points. Similarly, we obtain that human-taker trades affect prices on average by 0.6 basis points, while computer-taker trades affect prices on average by 0.3 basis points. Interestingly, focusing in the disaggregate order flow decomposition, we observe that when humans trade with other humans they influence prices more than when they trade with computers, and when computers trade with other computers they influence prices less than when they trade with humans. Our interpretation is that computers provide liquidity more strategically than humans, so that the counterparty cannot affect prices as much.

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, 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 Trading volumes in the dollar-swiss franc and euro-swiss franc markets tend to be close, with dollar-swiss franc volume a bit higher on average, but it is widely believed that price discovery occurs more often in the euro-swiss franc market than the dollar-swiss franc market. In this case HC order flow also has a slightly larger price impact in dollar-swiss franc than in euro-swiss franc.

The dynamics of the VAR system help reveal an interesting finding aside from the level of the price impact of order flow: There is a consistent and often large short-run over-reaction to CC shocks. That is, as seen both in Tables 6 through 9 and Figures 4 through 7, the short run response to a CC 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 shock is a 21 basis point move, but the long-run cumulative reaction is just 4 basis points (Table 8). Interestingly, the opposite pattern is true for the HH order flow shocks, where there is almost always an initial *under*-reaction in returns. To the extent that exchange rates follow random walks over medium term horizons, these impulse response patterns thus suggest that CC trading might contribute to excess short-run noise or volatility. This *over*-reaction disappears when we consider only human-taker and computer-taker order flow, but it is still significant in the euro-dollar and dollar-yen market when we consider the HH and computer-participation (at least one computer counterparty) order flow decomposition. One possible interpretation could be that the participation of computers in these markets, in whatever form, generates some excess short-run volatility.

In addition to the impulse response functions, we also report the long-run forecast variance decomposition of returns in Table 10 and 11 for the full sample and the three-month sub-sample, respectively.[13] That is, within the framework of the VAR, what fraction of the total (long-run) variance in returns can be attributed to innovations in the different order flows. As originally suggested by Hasbrouck (1991b), this variance decomposition can be interpreted 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 impulse response functions to one standard deviation shock to order flow, there are obvious patterns in the variance decompositions. The HH order flow makes up the dominant part of the variance share in most cases, which is not surprising given that this component constitutes the largest share on average across the sample period. 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 Figure 3. The mixed order flow (CH and HC order flow) typically contribute with about the same share to the explained variance. 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.

The seemingly disproportionately small fraction of the explained return variance that can be attributed to the CC order flow is likely a result both of the generally smaller responses by returns to shocks from this order flow component, as seen in the impulse response analysis, 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, presented in Table 1.[14,15] However, the (cumulative) impulse response functions to one-billion base currency shocks suggest that it is more due to the latter than to the former.

4.3 Summary

Our empirical analysis provides three important insights. First, we find that human order flow accounts for most of the (long-run) variance in exchange rate returns, i.e., humans are still the "informed" traders

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

[14] 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.

[15] 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.

in these markets. This can probably be attributed in part to the fact that investors are more likely to use algorithmic trading, relative to human trading, for the optimal execution of large orders at a minimum cost. Algorithmic trades appear to be successful at that task, so that computers break up the orders so as to have a small impact on prices.

Second, 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.[16]

Third, we show that there is an initial under-reaction to human-maker/human-taker order flow, while there is an initial over-reaction to computer-maker/computer-taker order flow. To the extent that there is an initial over-reaction to computer-maker/computer-taker order flow, we conclude that algorithmic trading may lead to some excess short-run volatility. There is also some evidence of over-reaction to order flow when computers participate in the market either as liquidity providers or liquidity demanders, but only for the euro-dollar and the dollar-yen markets.

5 Who provides liquidity during volatile times?

In the previous section 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. To further investigate this conjecture we estimate whether computer or humans are more likely to provide liquidity during volatile times. To that end we estimate the following system of equations,

$$\begin{aligned}\log V_{it}^{hmake} \quad &= \quad \sum_{h=1}^{48} \alpha_h D_h\left(t\right) + \sum_{j=1}^{q} \beta_{ij}^{r} \log\left|r_{it-j}\right| + \sum_{j=1}^{q} \beta_{ij}^{V^{hmake}} \log V_{it-j}^{hmake} \\ &\quad + \sum_{j=1}^{q} \beta_{ij}^{V^{cmake}} \log V_{it-j}^{cmake} + \sum_{k=1}^{K} \beta_k^{A} PA_{kt} + \beta_i^{RV} RV_{it} + u_{it}, \end{aligned} \tag{3}$$

[16] For example, Chakravarty and Holden (1995), Kumar and Seppi (1994), Kaniel and Liu (2006), 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.

$$\log V_{it}^{cmake} = \sum_{h=1}^{48} \alpha_h D_h(t) + \sum_{j=1}^{q} \beta_{ij}^{r} \log |r_{it-j}| + \sum_{j=1}^{q} \beta_{ij}^{V^{hmake}} \log V_{it-j}^{hmake} + \sum_{j=1}^{q} \beta_{ij}^{V^{cmake}} \log V_{it-j}^{cmake} + \sum_{k=1}^{K} \beta_k^A PA_{kt} + \beta_i^{RV} RV_{it} + u_{it}, \tag{4}$$

where $D_h(\cdot)$ is an indicator variable which takes on the value one if the observation at time t falls in hth half-hour slot of the day, V_{it}^{hmake} is the human-maker (HH+HC) trading volume of currency i at time t, V_{it}^{cmake} is the computer-maker (CC+CH) trading volume of currency i at time t, PA_{kt} captures the news effect on volatility, and RV_{it} is the daily realized volatility of currency i for the day on which observation t occurs. Our specification is similar to the specification in Andersen and Bollerslev (1998), who model intra-day volatility of the deutsche mark- dollar exchange rate.[17] That is, the log h-maker and h-taker volume is regressed on past values of volatility and past values of volume.[18] In addition, we also control for intra-daily seasonality in volume by including a set of dummy variables for each half-hour of the day; we also control for the overall level of the daily volatility, and thus the time trend in volatility and volume, by including the daily realized volatility calculated as the sum of squared one-minute returns; and we control for intra-day news effects. To promote tractability while at the same time maintaining flexibility, we impose a polynomial structure on the response patterns associated with news announcement k, PA_{kt}, similar to that adopted in Andersen et al. (2003). We allow the effect of news on volume to last for one hour after the announcement and we allow this effect to be different across news.

The estimation period is 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 sample. We use $q = 30$ lags of log volume and log exchange rate volatility in the estimation of equations (3) and (4). The main summary statistic we use in the presentation of the empirical results is the sum of the coefficients on the lagged volume and volatility. We therefore try not to include too many insignificant lagged volume terms, which would render the sums of the coefficients less informative. Since there were few significant coefficients at lags higher than 30, we set $q = 30$ in equations (3) and (4). The results are not particularly sensitive to the number of lags included, and our conclusions are qualitatively the same when we set $q = 5, 10$ and 20.

Our specification does not preclude endogeneity problems in the identification of the relationship, this is why we also consider an instrumental variables approach. However, this approach alleviates the problem to

[17] We regress the natural log of volume on the natural log of the absolute value of the exchange rate, rather than volume on the absolute value of the exchange rate because the residual properties of the former equation specification are nicer than those of the latter. In particular the residual of the log-specification is closer to a normal distribution than the residual of the levels-specification. We note, though, that our conclusions are robust to boh specification choices.

[18] We obtain similar results when we replace volume with the absolute value of order flow. The R-squared is slightly higher with the specification we report here, but our conclusions are qualitatively the same.

some extent and its simplicity makes it appealing. By including a large set of past own lags, along with the additional control variables, equations (3) and (4) attempt to control for possible correlation between the contemporaneous error term and the lagged volatility variable. Thus, although there is no guarantee that the volatility variables are exogenous in an econometric sense, most of the observed relationship between past volatility and future volume should be causal. In any case, equations (3) and (4) evaluate whether past volatility helps predict future volume, which is certainly of individual interest.

[TO BE COMPLETED]

6 Who demands liquidity and who provides liquidity during U.S. Macroeconomic News Announcements?

[TO BE COMPLETED]

7 How Correlated Are Algorithmic Trades and Strategies?

We investigate the proposition that AT agents tend to have trading strategies that are more correlated than those of human agents. 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.

If AT (computer) agents and human agents trade randomly, then we should expect to see them trading with each other in proportion to their relative presence in the market. If, on the other hand, computer agents tend to have more homogeneous trading strategies, we should expect to see them trading less among themselves and more with human agents. At the extreme, if all computer agents used the very same algorithms and had the exact same speed of execution, we would expect to see no trading volume among computers. Therefore, the fraction of trades conducted between computers agents contains information on how correlated their strategies are. To test this question, we assume a simple market model in which computer agents and human agents trade randomly, and then compare the implications of that model to the actual data.

In this model, there are two separate types of agents: makers and takers. Within each of these groups, there are both computer agents and human agents. During any given period k, computer agents make up some fixed proportion $\alpha_{m,k}$ of makers and some fixed proportion $\alpha_{t,k}$ of takers. We allow these proportions to differ from one another and to vary between periods. The remaining makers and takers are human agents, in proportions $(1 - \alpha_{m,k})$ and $(1 - \alpha_{t,k})$, respectively. The model abstracts from the fact that, in practice, actual traders can act as both makers and takers.

At each time k, we allow the agents to trade. We assume that any order submitted by a taker will be randomly matched with a maker, regardless of the identity of either party as a computer or a human. Thus, trading among computer agents, among human agents, and between computer agents and human agents should occur in proportion to those agents' relative presence on either side of the market.

We use the following nomenclature to refer to the four types of trading volume during period k: $VolHH_k$ refers to trading volume between human makers and human takers, $VolHC_k$ to trading volume between human makers and computer takers, $VolCH_k$ to trading volume between computer makers and human takers, and $VolCC_k$ to trading volume between computer makers and computer takers. Similarly, $PctHH_k$, $PctHC_k$, $PctCH_k$ and $PctCC_k$ refer to those volumes expressed as a percent of total trading volume. Under our random trading model, the following conditions hold:

$PctHH_k = (1 - \alpha_{m,k}) * (1 - \alpha_{t,k})$

$PctHC_k = (1 - \alpha_{m,k}) * \alpha_{t,k}$

$PctCH_k = \alpha_{m,k} * (1 - \alpha_{t,k})$

$PctCC_k = \alpha_{m,k} * \alpha_{t,k}$

Since $\alpha_{m,k}$ and $\alpha_{t,k}$ may vary over time, the model provides no guidance as to the expected levels of $PctHH_k$, $PctHC_k$, $PctCH_k$ and $PctCC_k$. But using the assumption of random trading between AT agents and human agents, the model does provide some guidance as to the proportions of these quantities. In particular, the model implies that

$\frac{VolHH_k}{VolCH_k} = \frac{VolHC_k}{VolCC_k}$,

for all strictly positive values of $\alpha_{m,k}$ and $\alpha_{t,k}$. For simplicity of notation, we define two key measures:

$RH_k = \frac{VolHH_k}{VolCH_k}$, the "human taker ratio"

$RC_k = \frac{VolHC_k}{VolCC_k}$, the "computer taker ratio"

Intuitively, for each agent type, the taker ratio is an expression of the propensity of takers of that type to trade with human makers relative to computer makers. In a market with mostly human makers, we would expect these ratios to exceed 1, while in a market with mostly computer makers, we would expect these ratios to be smaller than 1.[19]

The EBS data allow us to calculate actual taker ratios for each trading day. Figure 8 below shows 50-day moving averages of the ratio of these ratios, $R_k = \frac{RC_k}{RH_k}$, for each currency pair from 2006 to 2007. Two aspects of the data emerge from these figures. First, for all currency pairs, both taker ratios exhibit a

[19] The use of this model, and the appeal to these ratios, is not an attempt to complicate a simple analysis. We only know the volume of completed trades in each category (HH, CC, HC, CH) ex-post, but not the number of trades attempted by each type of trader. This limits the number of variables relative to the number of unknowns we seek to find, which allows us to make statements about ratios only. But , we can get the information we seek from these two ratios.

declining trend over time, corresponding to the general increase in the proportion of market makers that are computer agents. Second, for all currency pairs, RC_k appears to be consistently greater than RH_k, so the ratios remain above 1. In other words, computer takers trade disproportionately more with human makers than do human takers themselves. We believe this is evidence that algorithmic strategies tend to be highly correlated, certainly less diverse than the trading strategies used by human traders.

8 Conclusion

Using high-frequency trading data for five exchange rates over 2006 and 2007, we analyze the impact of the growth of algorithmic trading on the spot interdealer foreign exchange market. We use econometric techniques that take into account the obvious trends and likely endogeneity between realized volatility and the presence of algorithmic trading to analyze the relationship between these two variables. Using these techniques, we detect no systematic linkage, positive or negative, between realized volatility and the share of algorithmic trades. However, our instruments are weak, so we also analyze return-order flow dynamics in a high-frequency VAR framework in the tradition of Hasbrouck (1991a) to look at the link between algorithmic and non-algorithmic trades and volatility. We find that non-algorithmic trades account for a substantially larger share of the price movements than would be expected given the sizable fraction of algorithmic traders, i.e., non-algorithmic traders are still the "informed" traders in this market. We also find, in some cases, an initial over-reaction of the price to trades between algorithmic counterparties, which may be evidence that algorithmic trading can lead to some excess short-run volatility.

[to be completed]

References

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

[2] Andersen, T.G., 1996. Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility, *Journal of Finance* 51, 169-204.

[3] 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.

[4] Andersen, T.G., T. Bollerslev, and F.X. Diebold, 2007. Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility, *Review of Economics and Statistics* 89, 701-720.

[5] 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.

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

[7] 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.

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

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

[10] 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.

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

[12] 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.

[13] 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.

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

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

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

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

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

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

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

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

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

[23] Hansen, P.R., and A. Lunde, 2006. Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127–161.

[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 analsysis 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] 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, values 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. H-maker and H-taker represents a human maker and taker, respectively, and C-maker and C-taker represents a computer maker and taker. The definition of the different order flows are given in the main text, and the variable labelled total order flow is simply the sum of the four individual order flows. 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
			EUR/USD			
Returns	0.003	1.2398	-0.005^{***}	0.008	1.2057	0.007^{**}
Total order flow	0.0315	25.9455	0.150^{***}	-0.0937	29.7065	0.174^{***}
H-maker/H-taker	0.1425	19.9614	0.177^{***}	0.0327	21.9211	0.209^{***}
C-maker/H-taker	-0.1012	8.897	0.166^{***}	-0.1123	10.7649	0.215^{***}
H-maker/C-taker	0.0123	8.9232	0.152^{***}	0.0483	11.5856	0.150^{***}
C-maker/C-taker	-0.0222	2.7939	0.053^{***}	-0.0623	3.9477	0.072^{***}
			USD/JPY			
Returns	-0.0007	1.6038	-0.010^{***}	-0.0045	1.911	0.007^{**}
Total order flow	0.1061	20.098	0.189^{***}	-0.3439	23.6359	0.211^{***}
H-maker/H-taker	0.1037	15.9972	0.209^{***}	-0.1203	17.4612	0.226^{***}
C-maker/H-taker	-0.0184	6.903	0.172^{***}	-0.0885	9.1773	0.162^{***}
H-maker/C-taker	0.0198	7.5686	0.198^{***}	-0.0901	10.1673	0.191^{***}
C-maker/C-taker	0.0011	2.4556	0.032^{***}	-0.045	3.8751	0.026^{***}
			USD/CHF			
Returns	-0.0022	1.5856	-0.080^{***}	-0.0067	1.5473	-0.068^{***}
Total order flow	0.0159	7.3238	0.090^{***}	0.0071	8.2527	0.067^{***}
H-maker/H-taker	0.0173	3.8984	0.134^{***}	0.0073	3.7655	0.102^{***}
C-maker/H-taker	-0.0135	4.3146	0.152^{***}	-0.0327	5.5544	0.169^{***}
H-maker/C-taker	0.0188	4.4197	0.158^{***}	0.0568	5.4276	0.184^{***}
C-maker/C-taker	-0.0066	2.1487	0.007^{***}	-0.0243	2.9157	0.000
			EUR/CHF			
Returns	0.0009	0.7280	-0.101^{***}	0.0006	0.9023	-0.096^{***}
Total order flow	0.0043	6.1525	0.155^{***}	-0.0767	6.3677	0.160^{***}
H-maker/H-taker	0.0051	4.4847	0.116^{***}	-0.037	4.1587	0.128^{***}
C-maker/H-taker	-0.0173	2.0078	0.178^{***}	-0.0258	2.3073	0.166^{***}
H-maker/C-taker	0.0229	3.7705	0.272^{***}	-0.0037	4.3819	0.277^{***}
C-maker/C-taker	-0.0064	0.9417	0.058^{***}	-0.0102	1.295	0.049^{***}
			EUR/JPY			
Returns	0.0024	1.5976	-0.053^{***}	0.0036	2.1398	-0.017^{***}
Total order flow	-0.0648	7.0941	0.152^{***}	-0.1574	8.5978	0.147^{***}
H-maker/H-taker	-0.0172	4.4203	0.159^{***}	-0.06	4.3106	0.157^{***}
C-maker/H-taker	-0.0325	2.8912	0.129^{***}	-0.0616	3.7197	0.092^{***}
H-maker/C-taker	-0.0095	4.5331	0.173^{***}	-0.0264	6.0968	0.161^{***}
C-maker/C-taker	-0.0056	1.5558	0.023^{***}	-0.0095	2.5621	-0.001

Table 2: Summary statistics for the one-minute return and order flow data. The mean and standard deviation, as well as the first order autocorrelation, ρ, values 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. H-taker is the sum of human-maker/human-taker order flow plus computer-maker/human-taker order flow, C-taker is the sum of computer-maker/computer-taker order flow plus human-maker/computer-taker order flow, and C-involvement is is the sum of computer-maker/computer-taker order flow plus computer-maker/human-taker order flow, plus human-maker/computer-taker order flow. The definition of the different order flows are given in the main text. 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
			EUR/USD			
H-taker	0.0413	23.977	0.155***	−0.0796	26.8096	0.189***
C-taker	−0.0099	9.9363	0.127***	−0.014	12.89	0.115***
H-maker/H-taker	0.1425	19.9614	0.177***	0.0327	21.9211	0.209***
C-participation	−0.111	11.0735	0.061***	−0.1263	13.791	0.072***
			USD/JPY			
H-taker	0.0853	19.1127	0.190***	−0.2088	22.0344	0.204***
C-taker	0.0209	8.3941	0.170***	−0.1351	11.5877	0.158***
H-maker/H-taker	0.1037	15.9972	0.209***	−0.1203	17.4612	0.226***
C-participation	0.0025	8.6875	0.078***	−0.2236	11.7718	0.088***
			USD/CHF			
H-taker	0.0037	6.4585	0.148***	−0.0254	7.556	0.159***
C-taker	0.0122	5.0061	0.099***	0.0325	6.2702	0.110***
H-maker/H-taker	0.0173	3.8984	0.134***	0.0073	3.7655	0.102***
C-participation	−0.0013	5.3493	0.030***	−0.0002	6.5454	0.024***
			EUR/CHF			
H-taker	−0.0122	5.1276	0.109***	−0.0628	5.0371	0.114***
C-taker	0.0165	3.998	0.262***	−0.0139	4.7275	0.262***
H-maker/H-taker	0.0051	4.4847	0.116***	−0.037	4.1587	0.128***
C-participation	−0.0008	3.7605	0.173***	−0.0397	4.3822	0.162***
			EUR/JPY			
H-taker	−0.0497	5.7006	0.150***	−0.1216	6.2074	0.125***
C-taker	−0.0151	4.8409	0.146***	−0.0358	6.7	0.131***
H-maker/C-taker	−0.0095	4.5331	0.173***	−0.0264	6.0968	0.161***
C-involvement	−0.0476	4.6833	0.082***	−0.0974	6.4696	0.081***

Table 3: Correlation matrices for the one-minute return and order flow data. The correlations 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. H-maker and H-taker represents a human maker and taker, respectively, and C-maker and C-taker represents a computer maker and taker. The definition of the different order flows are given in the main text, and the variable labelled total order flow is simply the sum of the four individual order flows. There are a total of 717,120 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				
	Returns	Total order flow	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	Returns	Total order flow	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker
EUR/USD										
Total order flow	0.5789					0.5748				
H-maker/H-taker	0.5253	0.9162				0.4895	0.9005			
C-maker/H-taker	0.2639	0.4339	0.2739			0.2613	0.4101	0.2592		
H-maker/C-taker	0.2131	0.3754	0.1603	-0.3391		0.2711	0.4236	0.1924	-0.3647	
C-maker/C-taker	0.1016	0.1600	-0.0205	-0.0285	0.2267	0.0988	0.1633	-0.0481	-0.0103	0.1786
USD/JPY										
Total order flow	0.5573					0.5665				
H-maker/H-taker	0.4940	0.9088				0.4686	0.8783			
C-maker/H-taker	0.2884	0.4123	0.2795			0.3065	0.4262	0.3008		
H-maker/C-taker	0.1361	0.3072	0.0533	-0.3939		0.1767	0.3564	0.0633	-0.3956	
C-maker/C-taker	0.1128	0.1584	-0.0266	-0.0437	0.1923	0.1546	0.1972	-0.0273	-0.0856	0.2016
USD/CHF										
Total order flow	0.3955					0.4065				
H-maker/H-taker	0.3058	0.7044				0.2943	0.6346			
C-maker/H-taker	0.2487	0.4758	0.2348			0.2887	0.5066	0.2884		
H-maker/C-taker	0.0683	0.4083	0.0583	-0.4078		0.0290	0.3597	-0.0142	-0.4555	
C-maker/C-taker	0.1534	0.3353	-0.0049	0.0265	0.048	0.1666	0.3761	-0.0181	0.0044	0.0428
EUR/CHF										
Total order flow	0.4026					0.4055				
H-maker/H-taker	0.3312	0.7941				0.3243	0.7295			
C-maker/H-taker	0.0876	0.1755	0.1193			0.1014	0.1678	0.1435		
H-maker/C-taker	0.1977	0.5384	0.0462	-0.3895		0.2018	0.5895	0.0393	-0.4159	
C-maker/C-taker	0.0749	0.2216	-0.0131	0.0062	0.1241	0.0888	0.2806	-0.0128	-0.0101	0.1297
EUR/JPY										
Total order flow	0.4337					0.4685				
H-maker/H-taker	0.3292	0.7643				0.3496	0.6833			
C-maker/H-taker	0.1140	0.2793	0.1799			0.1185	0.2645	0.1907		
H-maker/C-taker	0.2471	0.5621	0.1100	-0.3746		0.2937	0.6426	0.1489	-0.3518	
C-maker/C-taker	0.1102	0.2316	-0.0112	-0.0044	0.0329	0.1129	0.2931	-0.0206	-0.0481	0.0370

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. Newey-West standard errors are given in parentheses below the coefficient estimates, calculated using 22 lags, and * indicates significance at the 10% level, ** significance at 5% level, *** significance at 1% level. The left hand side of the table shows the results without time dummies included in the regressions and the right hand side of the table shows the results with monthly time dummies included in the regressions. In addition to the fraction of algorithmic trading and the potential monthly time dummies, the lag structure described in the main text is also included in every specification. In all cases, only the coefficient on the fraction of algorithmic trading is displayed. Panel A shows the results from a standard OLS estimation, along with the R^2. Panel B shows the results from the IV specification using all five currency pairs, described in detail in the main text, and estimated with Limited Information Maximum Likelihood (LIML). Panel C shows the corresponding LIML results for the IV specification using just the three main currency pairs. In Panels B and C, the $J-$test of overidentifying restrictions, with the p-value in parentheses below, and 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 the five currency specification, in order for the actual size of the LIML test to be no greater than 10%, the F-statistic should exceed 5.44, for a size no greater than 15%, the F-statistic should exceed 3.87, and for 20% and 25%, the corresponding numbers are 3.30 and 2.98, respectively. The corresponding critical values in the three currency case are 8.68, 5.33, 4.42, and 3.92, for maximal LIML sizes of 10%, 15%, 20%, and 25%, respectively.

	Without time dummies					With time dummies				
	EUR/USD	USD/JPY	USD/CHF	EUR/CHF	EUR/JPY	EUR/USD	USD/JPY	USD/CHF	EUR/CHF	EUR/JPY
Panel A. OLS										
Coeff. on AT	0.0059***	0.002	0.0077***	0.0048***	0.0045***	0.0141***	0.0046	0.0183***	0.0069***	0.0091***
	(0.0022)	(0.0018)	(0.0023)	(0.0013)	(0.0012)	(0.0029)	(0.0031)	(0.0031)	(0.0017)	(0.0019)
R^2 (%)	46.38	56.37	42.24	58.27	69.85	54.76	62.09	52.91	65.09	74.94
Panel B. IV, five currencies										
Coeff. on AT	0.0015	−0.0025	0.001	0.0047***	−0.0009	0.0148	−0.3064	0.0639	0.0589	0.0603
	(0.0036)	(0.0024)	(0.0033)	(0.0017)	(0.0018)	(0.0216)	(6.2268)	(0.0516)	(0.0635)	(0.0751)
J-Stat	9.57	8.45	8.44	3.31	8.49	6.87	0.31	3.19	0.47	1.46
P-Val	0.0226	0.0376	0.0377	0.3457	0.0369	0.0763	0.9572	0.3637	0.9250	0.6916
F-Stat	30.74	41.40	30.41	31.35	30.24	6.64	2.40	3.57	0.63	1.35
Panel C. IV, three currencies										
Coeff. on AT	−0.0009	−0.0003	0.0003			0.0041	0.0039	0.0802		
	(0.0015)	(0.0014)	(0.0017)			(0.0137)	(0.0340)	(0.0542)		
J-Stat	3.71	0.13	4.82			0.63	1.52	0.22		
P-Val	0.0540	0.7178	0.0281			0.4279	0.2184	0.6370		
F-Stat	312.46	118.32	51.54			8.75	3.74	2.06		

Table 5: In this table we report selected coefficient estimates of equation (1). We also report the R-squared of the full specification with a total of 717,120 observations, and the R-squared in a regression that only uses macroeconomic news announcements as the independent regressors, this specification only uses those observations when there was an announcement and in total there are 441 observations. The ***, **, and * represents significance at the 1, 5, and 10 percent level, respectively

	EUR/USD	USD/JPY	USD/CHF	EUR/CHF	EUR/JPY
	Quarterly Announcements				
1- GDP Advance	−11.744***	8.764***	11.136***	0.278	−3.439***
2- GDP Preliminary	−6.421***	9.282***	7.703***	0.154	−0.355
3- GDP Final	−0.281	1.319	2.320**	−0.739*	0.720
	Monthly Announcements				
Real Activity					
4- Unemployment Rate	6.125***	−14.209***	−16.223***	−2.277***	−7.355***
5- Nonfarm Payroll Employment	−40.398***	37.04***	47.728***	7.514***	9.477***
6- Retail Sales	−6.044***	4.059***	7.125***	0.271*	−0.877***
7- Industrial Production	−0.291	0.621*	1.244***	0.190	0.034
8- Capacity Utilization	−0.498	1.485***	0.964**	0.023	0.106
9- Personal Income	−0.418*	3.800***	1.012***	−0.127	−0.127
10- Consumer Credit	0.206	0.180	0.361	0.802***	0.385
Consumption					
11- New Home Sales	−2.546***	3.448***	4.596***	0.525***	0.993***
12- Personal Consumption Exp.	−2.215***	0.751**	1.463***	−0.712***	0.444
Investment					
13- Durable Goods Orders	−1.197***	1.703***	3.381***	0.279**	0.453
14- Construction Spending	−4.490***	4.71***	4.305***	−0.626***	−0.770
15- Factory Orders	−0.390*	0.362	1.327***	0.087	0.341
16- Business Inventories	0.998***	−0.505	1.582***	0.646***	−0.520
Government Purchases					
17- Government Budget	0.169	−0.241	−0.341	−0.02	−0.222
Net Exports					
18- Trade Balance	−6.234***	3.893***	5.716***	−0.326**	−0.652**
Prices					
19- Producer Price Index	0.181	−1.106***	0.381**	0.073	−0.615***
20- Core PPI	−3.511***	1.369***	2.892***	−0.22**	0.157
21- Consumer Price Index	−1.746***	0.729**	1.287***	0.29*	−0.508
22- Core CPI	−25.26***	12.864***	26.735***	−1.408***	−6.743***
Forward-Looking					
23- Consumer Confidence Index	−2.963***	4.944***	8.967***	0.635***	−0.62
24- NAPM Index	−3.327***	4.067***	7.587***	0.117	−0.458
25- Housing Starts	−2.486***	1.821***	4.838***	−0.082	−0.472
26- Index of Leading Indicators	1.103***	0.488	−0.674	−0.001	−0.286
	Six-Week Announcements				
27- Target Federal Funds Rate	−0.404	2.185***	6.458***	−4.128***	−11.129***
	Weely Announcements				
28- Initial Unemployment Claims	1.615***	−1.165***	−2.450***	0.060	0.523***
	Contemporaneous Order Flow				
H-maker/H-taker	27.636***	43.495***	98.998***	50.889***	102.628***
C-maker/H-taker	29.665***	58.951***	93.028***	47.616***	92.205***
H-maker/C-taker	26.576***	40.365***	50.218***	48.611***	100.916***
C-maker/C-taker	32.192***	61.615***	100.8***	37.140***	102.077***
	R-squared				
Total (717,120 observations)	35.65%	33.47%	19.39%	18.89%	20.34%
Announcements Only (441 obs)	59.96%	58.38%	56.02%	32.98%	22.65%

Table 6: Impulse responses from the VAR estimation using the entire 2006-2007 sample. Panels A1-A3 show the short-run (immediate) response of returns to one billion base currency shock to a particular type of order flow. The response is measured in basis points. Panels B1-B3 show the corresponding cumulative long-run response of returns, calculated as the cumulative impact of the shock after 30 minutes. Panels C1-C3 provide the difference between the cumulative long-run response in returns minus the immediate response of returns, i.e., it provides the extent of over-reaction or under-reaction to an order flow shock. Panels A1, B1, and C1, give the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by all four order flows, HH, CH, HC, and CC. Panels A2, B2, and C2, give the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by two order flows, H-taker (HH+CH) and C-maker (CC+HC). Panels A3, B3, and C3, give the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by two order flows, H-maker/H-taker (HH) and C-participation (CC+HC+CH).

	Four Order Flows				Who Initiates the Trade		Computer Participation	
	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker	H-taker	C-taker	H-maker/ H-taker	C-Participation
	Panel A1: SR impulse response (bp)				Panel A2: SR (bp)		Panel A3: SR (bp)	
EUR/USD	27.64	29.66	26.57	32.19	28.06	26.94	27.72	28.34
USD/JPY	43.48	58.94	40.34	61.57	46.77	39.81	44.60	48.89
USD/CHF	98.98	93.00	50.21	100.72	97.83	62.24	101.64	75.75
EUR/CHF	50.89	47.61	48.58	37.11	50.23	48.06	50.96	47.27
EUR/JPY	102.61	92.16	100.91	102.04	99.32	102.71	101.80	99.64
	Panel B1: LR impulse response (bp)				Panel B2: LR (bp)		Panel B3: LR (bp)	
EUR/USD	30.13	20.47	29.89	24.92	27.87	32.35	29.96	25.34
USD/JPY	47.01	49.53	42.61	54.37	47.50	44.27	47.40	46.38
USD/CHF	103.11	82.44	64.89	90.04	93.42	74.31	105.99	76.16
EUR/CHF	55.31	40.26	53.61	31.75	52.38	54.16	54.82	50.50
EUR/JPY	116.12	91.24	107.18	93.41	108.07	109.85	115.15	103.13
	Panel C1: Difference LR-SR				Panel C2: LR-SR		Panel C3: LR-SR	
EUR/USD	2.49	−9.19	3.32	−7.26	−0.20	5.42	2.24	−3.00
USD/JPY	3.53	−9.41	2.27	−7.20	0.74	4.46	2.80	−2.50
USD/CHF	4.13	−10.55	14.67	−10.68	−4.41	12.07	4.35	0.41
EUR/CHF	4.42	−7.35	5.03	−5.36	2.15	6.09	3.86	3.23
EUR/JPY	13.51	−0.92	6.27	−8.63	8.75	7.14	13.35	3.49

Table 7: Impulse responses from the VAR estimation using the entire 2006-2007 sample. Panels A1-A3 show the short-run (immediate) response of returns to one standard deviation base currency shock to a particular type of order flow. The response is measured in basis points. Panels B1-B3 show the corresponding cumulative long-run response of returns, calculated as the cumulative impact of the shock after 30 minutes. Panels C1-C3 provide the difference between the cumulative long-run response in returns minus the immediate response of returns, i.e., it provides the extent of over-reaction or under-reaction to an order flow shock. Panels A1, B1, and C1, give the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by all four order flows, HH, CH, HC, and CC. Panels A2, B2, and C2, give the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by two order flows, H-taker (HH+CH) and C-maker (CC+HC). Panels A3, B3, and C3, give the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by two order flows, H-maker/H-taker (HH) and C-participation (CC+HC+CH).

	Four Order Flows				Who Initiates the Trade		Computer Participation	
	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker	H-taker	C-taker	H-maker/ H-taker	C-Participation
	Panel A1: SR impulse response (bp)				Panel A2: SR (bp)		Panel A3: SR (bp)	
EUR/USD	0.5389	0.2575	0.2318	0.0893	0.6617	0.2639	0.5410	0.3117
USD/JPY	0.6721	0.3968	0.2962	0.1506	0.8706	0.3269	0.6899	0.4215
USD/CHF	0.3789	0.3946	0.2169	0.2154	0.6214	0.3073	0.3896	0.4042
EUR/CHF	0.2260	0.0927	0.1737	0.0348	0.2554	0.1831	0.2264	0.1736
EUR/JPY	0.4440	0.2629	0.4481	0.1583	0.5572	0.4901	0.4405	0.4640
	Panel B1: LR impulse response (bp)				Panel B2: LR (bp)		Panel B3: LR (bp)	
EUR/USD	0.5875	0.1777	0.2608	0.0692	0.6570	0.3170	0.5847	0.2787
USD/JPY	0.7267	0.3334	0.3129	0.1330	0.8843	0.3635	0.7332	0.4000
USD/CHF	0.3947	0.3498	0.2803	0.1926	0.5934	0.3668	0.4063	0.4064
EUR/CHF	0.2456	0.0784	0.1917	0.0298	0.2664	0.2063	0.2435	0.1855
EUR/JPY	0.5024	0.2603	0.4760	0.1449	0.6063	0.5242	0.4983	0.4803
	Panel C1: Difference LR-SR				Panel C2: LR-SR		Panel C3: LR-SR	
EUR/USD	0.0486	−0.0798	0.0290	−0.0202	−0.0046	0.0531	0.0437	−0.0330
USD/JPY	0.0546	−0.0634	0.0167	−0.0176	0.0137	0.0366	0.0432	−0.0216
USD/CHF	0.0158	−0.0448	0.0634	−0.0229	−0.0280	0.0596	0.0167	0.0022
EUR/CHF	0.0196	−0.0143	0.0180	−0.0050	0.0110	0.0232	0.0171	0.0119
EUR/JPY	0.0584	−0.0026	0.0279	−0.0134	0.0491	0.0341	0.0578	0.0163

Table 8: Impulse responses from the VAR estimation using the three-month sub-sample. Only observations from September, October, and November of 2007 are used in the analysis. Panels A1-A3 show the short-run (immediate) response of returns from a shock to a particular type of order flow. The response is measured in basis points, and the size of the shock is always one billion of order flow in the base currency. Panels B1-B3 show the corresponding cumulative long-run response of returns, calculated as the cumulative impact of the shock after 30 minutes. Panels C1-C3 provide the extent of over-reaction or under-reaction to a shock of one of the order flows, i.e., it provides the difference between the cumulative long-run response in returns minus the immediate response of returns. Panels A1, B1, and C1, give the results for the structural identification scheme of the VAR, where returns is affected contemporaneously by all four order flows, but there is no contemporaneous feedback between the order flow themselves or from returns to the any of the order flows. Panels A2, B2, and C2, give the results for the structural identification scheme of the VAR, where returns is affected contemporaneously by two order flows, H-taker (HH+CH) and C-maker (CC+HC) but there is no contemporaneous feedback between the order flow themselves or from returns to the any of the order flows. Panels A3, B3, and C3, give the results for the structural identification scheme of the VAR, where returns is affected contemporaneously by two order flows, H-maker/H-taker (HH) and C-participation (CC+HC+CH) but there is no contemporaneous feedback between the order flow themselves or from returns to the any of the order flows.

	Four Order Flows				Who Initiates the Trade		Computer Participation	
	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker	H-taker	C-taker	H-maker/ H-taker	C-Participation
	Panel A1: SR impulse response (bp)				Panel A2: SR (bp)		Panel A3: SR (bp)	
EUR/USD	20.58	30.94	28.94	21.74	23.20	25.22	21.08	28.37
USD/JPY	41.96	64.63	46.08	67.65	48.02	44.89	43.32	54.69
USD/CHF	87.66	86.49	45.88	84.74	89.24	56.20	94.59	69.68
EUR/CHF	65.25	59.69	54.27	40.34	63.66	53.47	65.90	52.86
EUR/JPY	139.33	103.92	114.01	94.47	124.02	115.52	139.30	109.23
	Panel B1: LR impulse response (bp)				Panel B2: LR (bp)		Panel B3: LR (bp)	
EUR/USD	24.18	23.35	34.64	5.94	24.16	31.38	25.25	25.96
USD/JPY	46.83	57.24	40.33	51.81	49.54	40.63	48.25	46.85
USD/CHF	94.14	76.64	54.31	78.41	84.59	63.04	99.05	68.32
EUR/CHF	62.17	54.75	61.66	34.30	60.13	59.46	62.44	57.47
EUR/JPY	159.46	96.85	118.47	95.20	132.53	123.26	157.99	111.89
	Panel C1: Variance decomposition (%)				Panel C2: LR-SR		Panel C3: LR-SR	
EUR/USD	3.60	−7.59	5.70	−15.80	0.96	6.16	4.17	−2.41
USD/JPY	4.87	−7.39	−5.75	−15.85	1.52	−4.26	4.93	−7.84
USD/CHF	6.47	−9.85	8.44	−6.33	−4.65	6.83	4.45	−1.36
EUR/CHF	−3.08	−4.94	7.39	−6.03	−3.53	6.00	−3.46	4.61
EUR/JPY	20.13	−7.07	4.46	0.74	8.51	7.74	18.69	2.66

Table 9: Impulse responses from the VAR estimation using the three-month sub-sample. Only observations from September, October, and November of 2007 are used in the analysis. Panels A1-A3 show the short-run (immediate) response of returns from one standard deviation shock to a particular type of order flow. The response is measured in basis points, and the size of the shock is always one standard deviation. Panels B1-B3 show the corresponding cumulative long-run response of returns, calculated as the cumulative impact of the shock after 30 minutes. Panels C1-C3 provide the extent of over-reaction or under-reaction to a shock of one of the order flows , i.e., it provides the difference between the cumulative long-run response in returns minus the immediate response of returns. Panels A1, B1, and C1, give the results for the structural identification scheme of the VAR, where returns is affected contemporaneously by all four order flows, but there is no contemporaneous feedback between the order flow themselves or from returns to the any of the order flows. Panels A2, B2, and C2, give the results for the structural identification scheme of the VAR, where returns is affected contemporaneously by two order flows, H-taker (HH+CH) and C-maker (CC+HC) but there is no contemporaneous feedback between the order flow themselves or from returns to the any of the order flows. Panels A3, B3, and C3, give the results for the structural identification scheme of the VAR, where returns is affected contemporaneously by two order flows, H-maker/H-taker (HH) and C-participation (CC+HC+CH) but there is no contemporaneous feedback between the order flow themselves or from returns to the any of the order flows.

	Four Order Flows				Who Initiates the Trade		Computer Participation	
	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker	H-taker	C-taker	H-maker/ H-taker	C-Participation
	Panel A1: SR impulse response (bp)				Panel A2: SR (bp)		Panel A3: SR (bp)	
EUR/USD	0.4342	0.3211	0.3228	0.0845	0.6045	0.3181	0.4455	0.3854
USD/JPY	0.7019	0.5801	0.4544	0.2607	1.0241	0.5098	0.7251	0.6383
USD/CHF	0.3252	0.4693	0.2410	0.2451	0.6603	0.3453	0.3512	0.4547
EUR/CHF	0.2675	0.1333	0.2240	0.0518	0.3171	0.2397	0.2703	0.2256
EUR/JPY	0.5859	0.3829	0.6809	0.2409	0.7587	0.7636	0.5862	0.7016
	Panel B1: LR impulse response (bp)				Panel B2: LR (bp)		Panel B3: LR (bp)	
EUR/USD	0.5101	0.2424	0.3864	0.0231	0.6296	0.3957	0.5335	0.3527
USD/JPY	0.7834	0.5137	0.3976	0.1997	1.0565	0.4614	0.8076	0.5468
USD/CHF	0.3492	0.4158	0.2853	0.2268	0.6259	0.3873	0.3677	0.4459
EUR/CHF	0.2549	0.1222	0.2545	0.0441	0.2995	0.2666	0.2562	0.2452
EUR/JPY	0.6706	0.3568	0.7076	0.2428	0.8108	0.8148	0.6648	0.7187
	Panel C1: Difference LR-SR				Panel C2: LR-SR		Panel C3: LR-SR	
EUR/USD	0.0760	−0.0788	0.0636	−0.0614	0.0251	0.0777	0.0880	−0.0327
USD/JPY	0.0815	−0.0663	−0.0567	−0.0611	0.0324	−0.0483	0.0825	−0.0915
USD/CHF	0.0240	−0.0535	0.0443	−0.0183	−0.0344	0.0420	0.0165	−0.0088
EUR/CHF	−0.0126	−0.0110	0.0305	−0.0077	−0.0176	0.0269	−0.0142	0.0197
EUR/JPY	0.0847	−0.0260	0.0266	0.0019	0.0520	0.0512	0.0787	0.0171

Table 10: Variance decompositions from the VAR estimation using the entire 2006-2007 sample. Panels A1-A3 provide the long-run variance decomposition of returns, expressed in percent and calculated at the 30 minute horizon. Panel A1 shows the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by all four order flows, HH, CH, HC, and CC. Panel A2 gives the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by two order flows, H-taker (HH+CH) and C-maker (CC+HC). Panel A3 gives the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by two order flows, H-maker/H-taker (HH) and C-participation (CC+HC+CH).

	Panel A1: Four Order Flows				Panel A2 : Who Initiates the Trade		Panel A3: Computer Participation	
	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker	H-taker	C-taker	H-maker/ H-taker	C-Participation
EUR/USD	20.71	4.73	3.89	0.58	29.27	4.74	21.30	7.07
USD/JPY	18.62	6.48	3.70	0.93	29.31	4.22	20.05	7.52
USD/CHF	5.89	6.39	1.95	1.90	15.22	3.74	6.36	6.84
EUR/CHF	9.67	1.63	5.74	0.23	12.16	6.29	9.88	5.83
EUR/JPY	7.84	2.74	7.94	0.99	12.03	9.28	7.98	8.79

Table 11: Variance decompositions from the VAR estimation using the three-month sub-sample. Panels A1-A3 provide the long-run variance decomposition of returns, expressed in percent and calculated at the 30 minute horizon. Panel A1 shows the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by all four order flows, HH, CH, HC, and CC. Panel A2 gives the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by two order flows, H-taker (HH+CH) and C-maker (CC+HC). Panel A3 gives the results for the structural identification scheme of the VAR, where returns are affected contemporaneously by two order flows, H-maker/H-taker (HH) and C-participation (CC+HC+CH). Only observations from September, October, and November of 2007 are used in the analysis.

	Panel A1 Four Order Flows				Panel A2 Who Initiates the Trade		Panel A3 Computer Participation	
	H-maker/ H-taker	C-maker/ H-taker	H-maker/ C-taker	C-maker/ C-taker	H-taker	C-taker	H-maker/ H-taker	C-Participation
EUR/USD	14.19	7.68	7.86	0.59	25.92	7.25	15.59	11.50
USD/JPY	14.47	9.78	6.12	2.00	28.59	7.22	16.10	12.46
USD/CHF	4.63	9.59	2.57	2.66	18.00	4.98	5.58	9.34
EUR/CHF	8.89	2.23	6.34	0.36	12.19	7.13	9.27	6.53
EUR/JPY	7.72	3.32	10.47	1.30	12.47	12.67	8.04	11.52

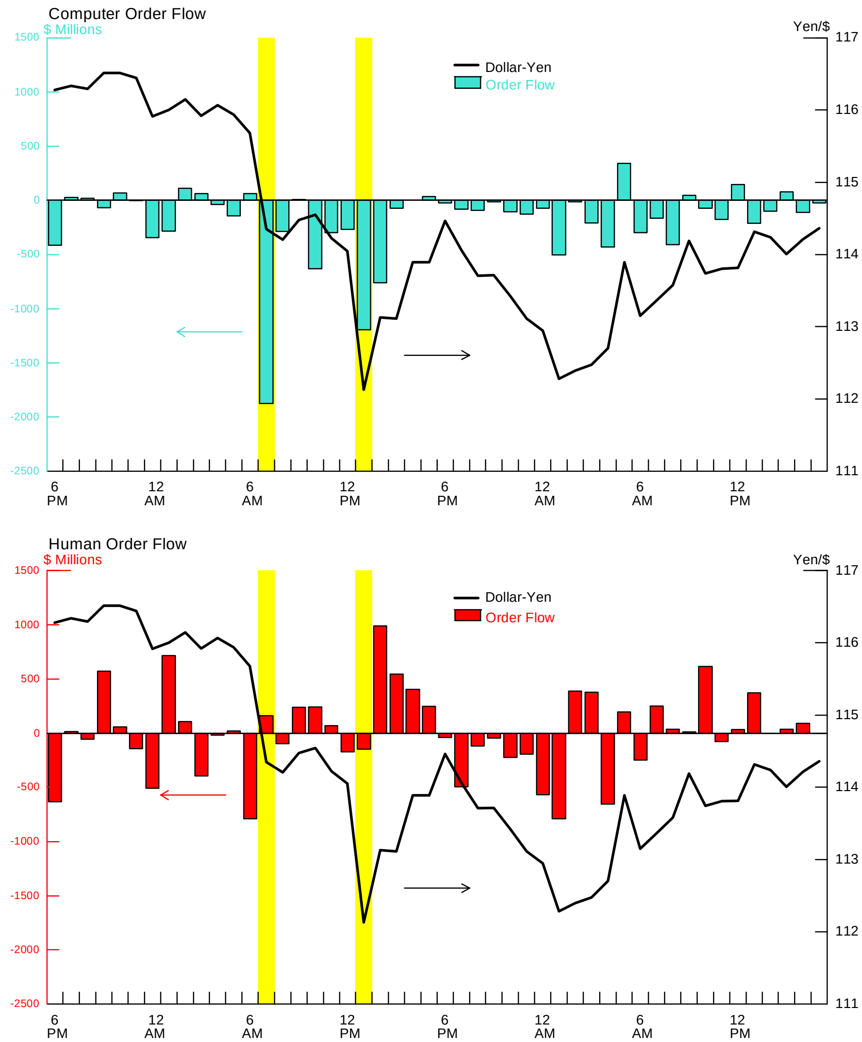

Figure 1: Dollar-Yen Market on August 16, 2007

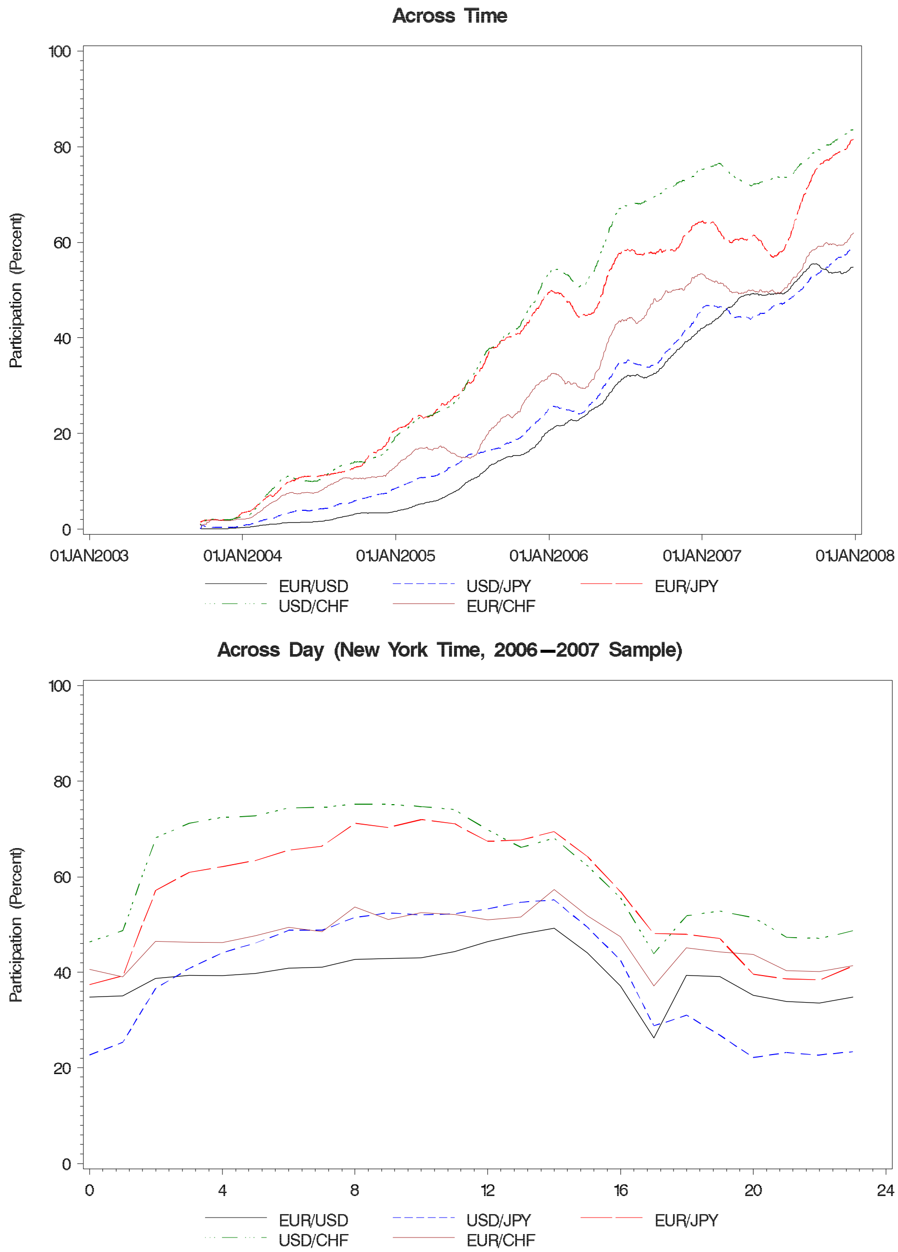

Figure 2: Participation rates of algorithmic traders.

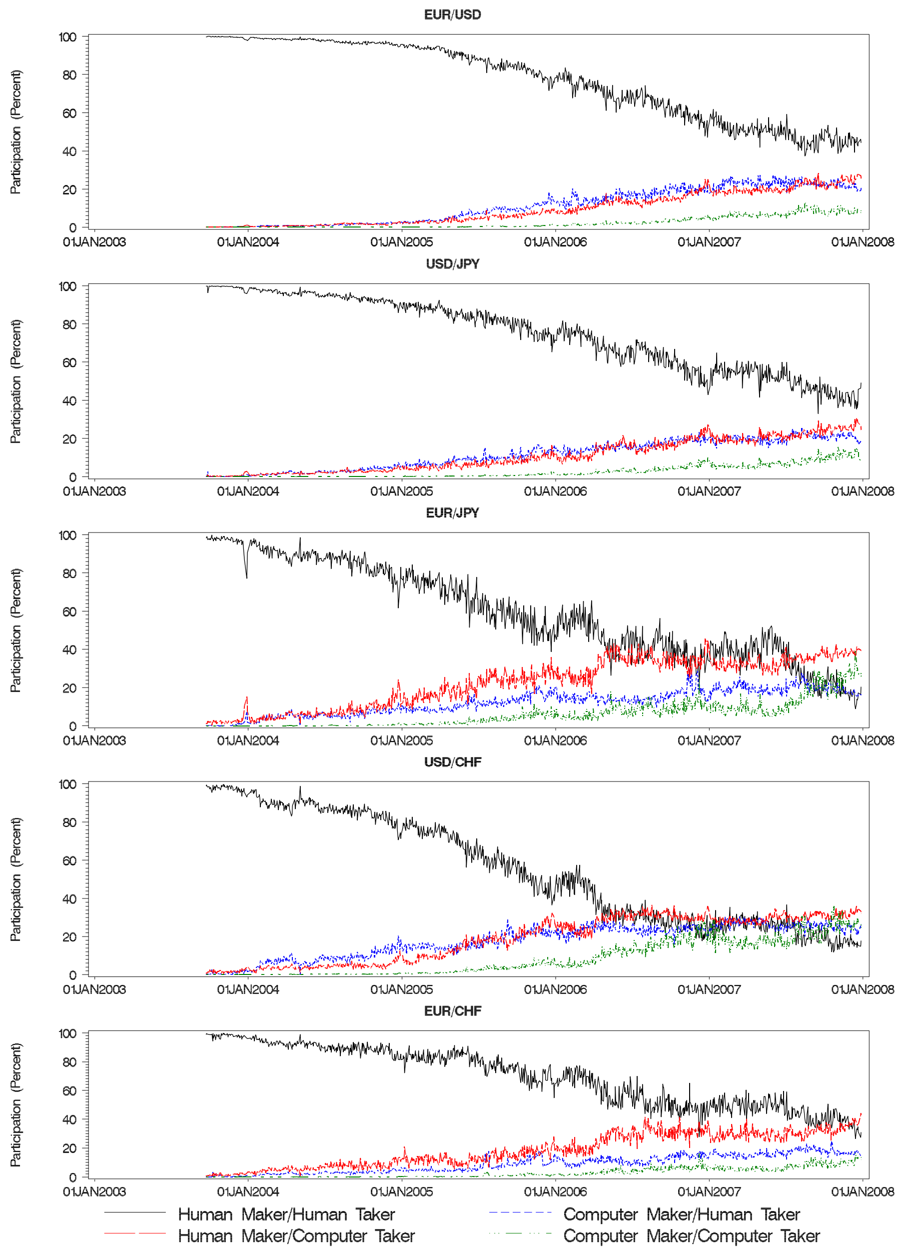

Figure 3: Participation rates of algorithmic traders broken down into categories.

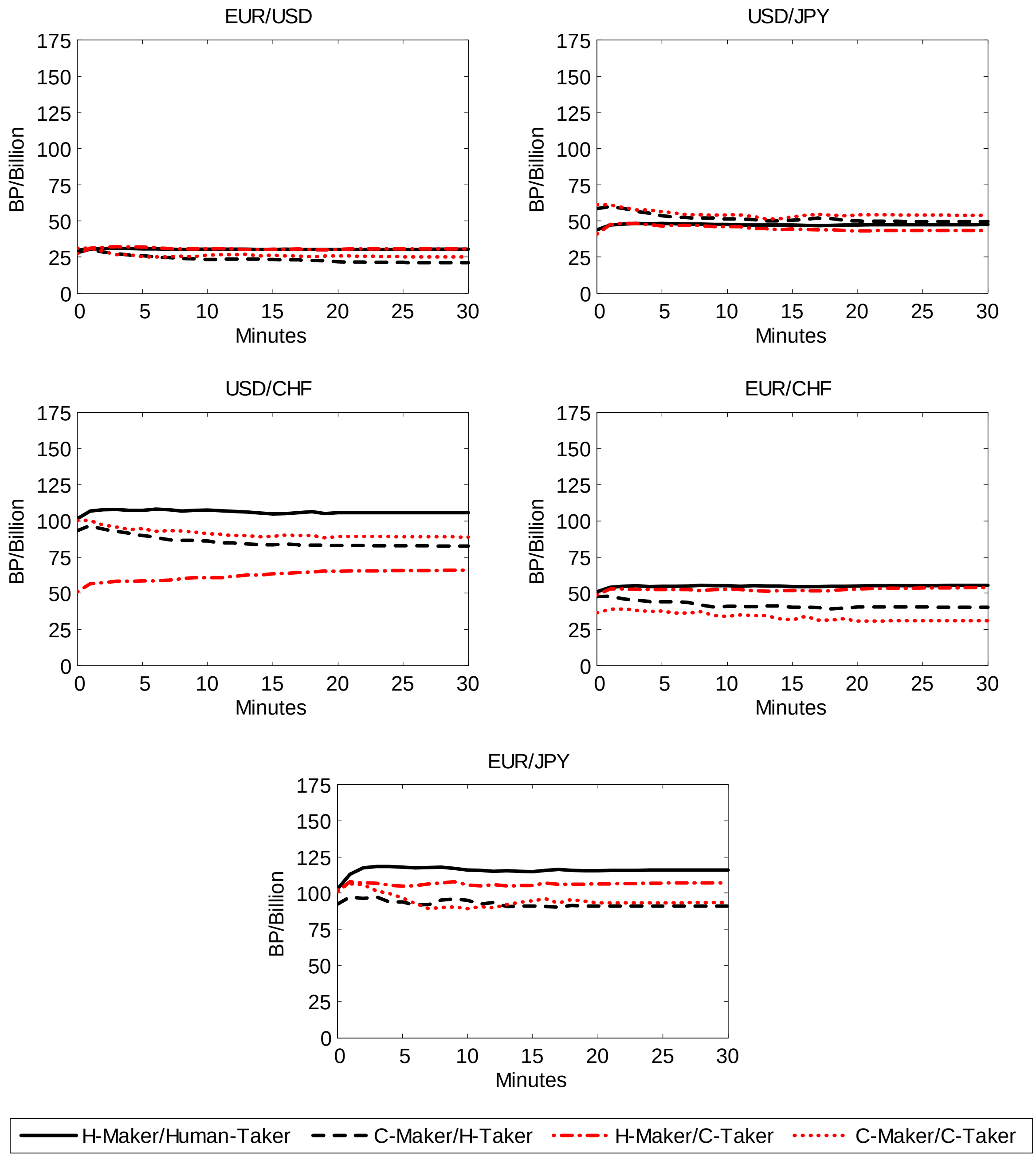

Figure 4: Cumulative impulse response functions for returns based on the entire 2006-2007 sample. The graphs trace out the cumulative impulse responses of returns from a shock to one of the four order flow components; H-maker and H-taker represents a human maker and taker, respectively, and C-maker and C-taker represents a computer maker and taker, respectively. The response is measured in basis points, and the size of the shock is always one billion of order flow in the base currency. The results are obtained under the structural identification scheme for the VAR, where returns is affected contemporaneously by all four order flows, but there is no contemporaneous feedback between the order flow themselves or from returns to the any of the order flows.

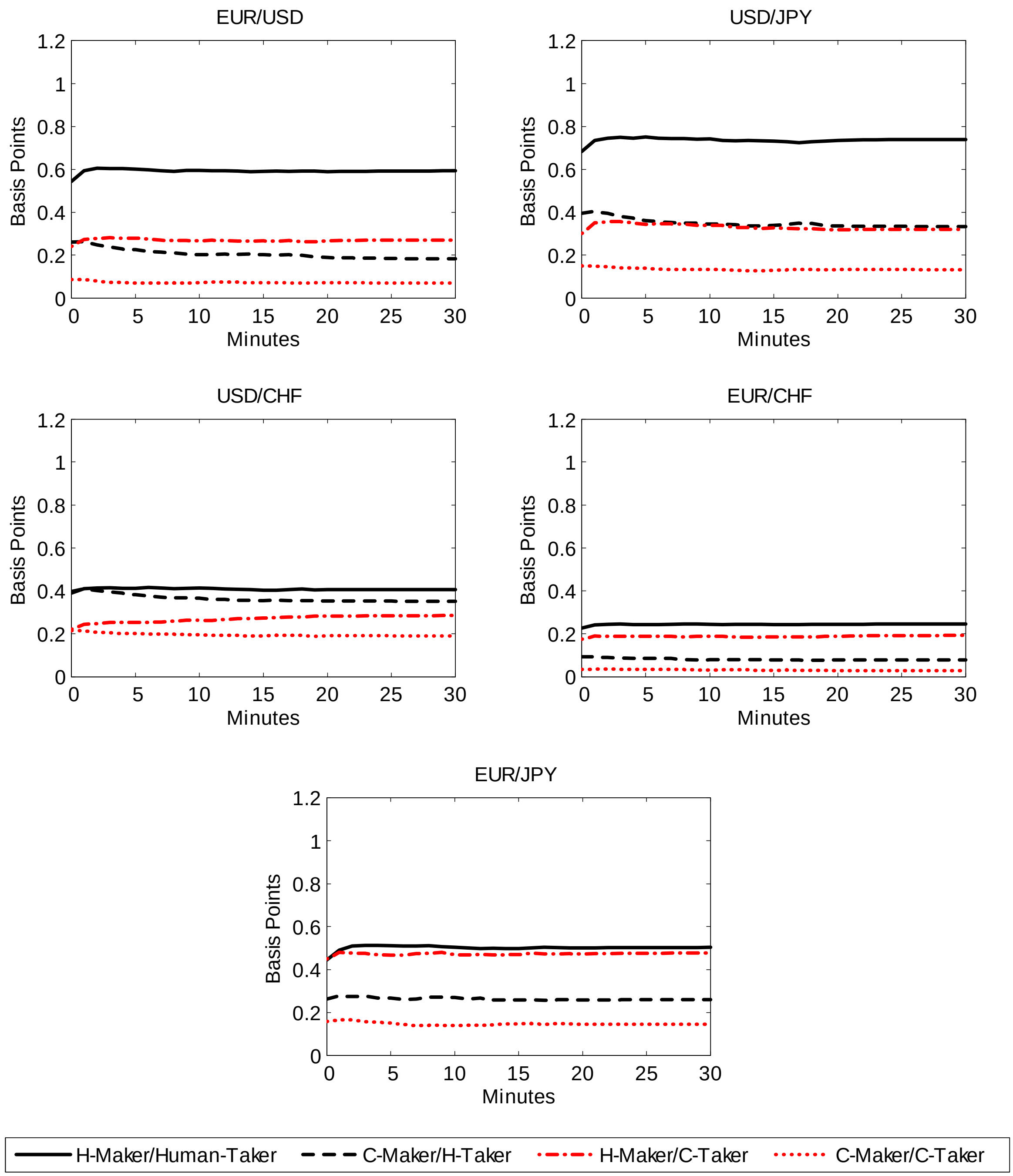

Figure 5: Cumulative impulse response functions to a standard deviation shock for returns based on the entire 2006-2007 sample.

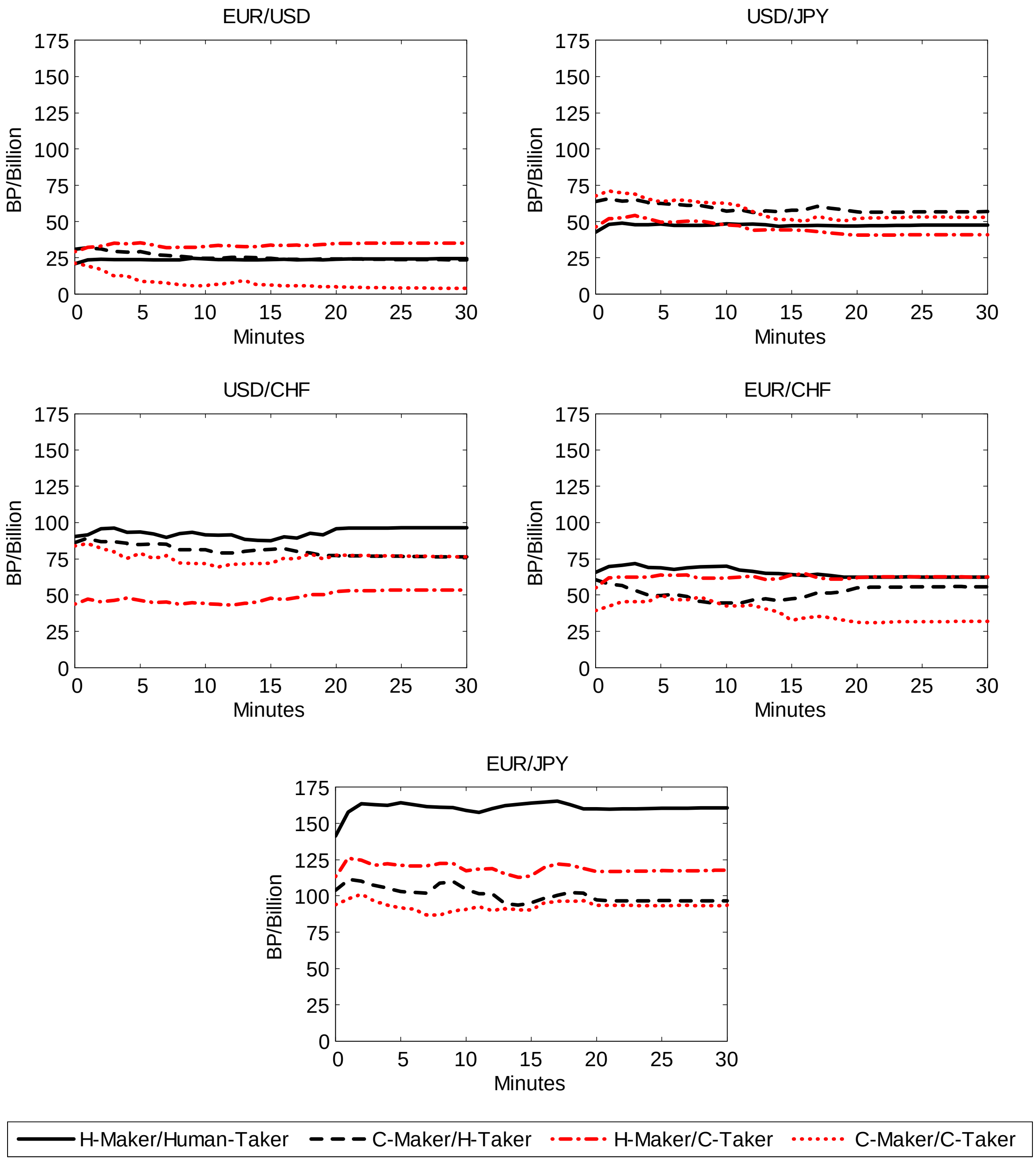

Figure 6: Cumulative impulse response functions for returns based on the three-month sub-sample. The graphs trace out the cumulative impulse responses of returns from a shock to one of the four order flow components; H-maker and H-taker represents a human maker and taker, respectively, and C-maker and C-taker represents a computer maker and taker, respectively. The response is measured in basis points, and the size of the shock is always one billion of order flow in the base currency. The results are obtained under the structural identification scheme for the VAR, where returns is affected contemporaneously by all four order flows, but there is no contemporaneous feedback between the order flow themselves or from returns to the any of the order flows. Only observations from September, October, and November of 2007 are used.

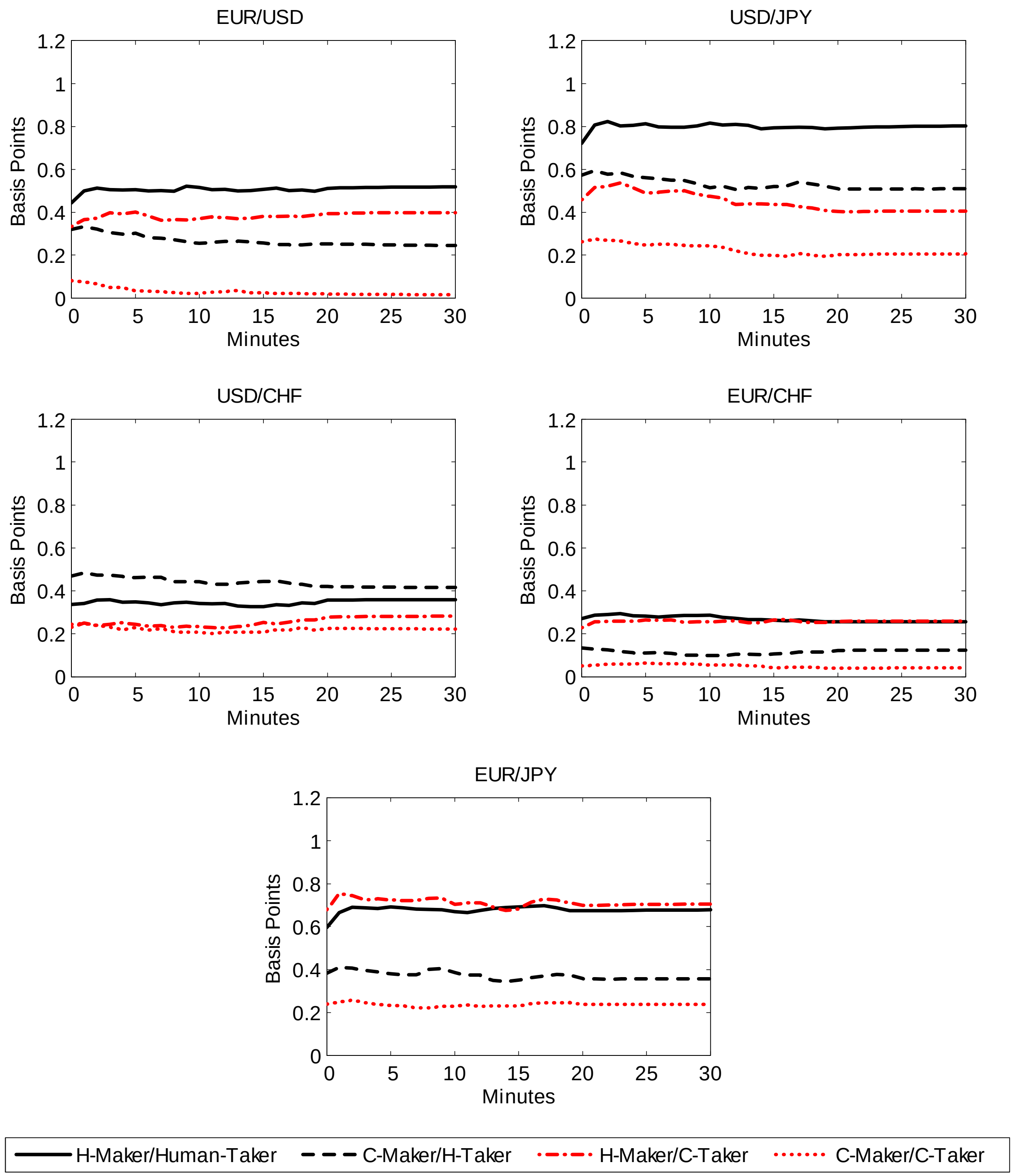

Figure 7: Cumulative impulse response functions to a standard deviation shock for returns based on the three-month sub-sample.

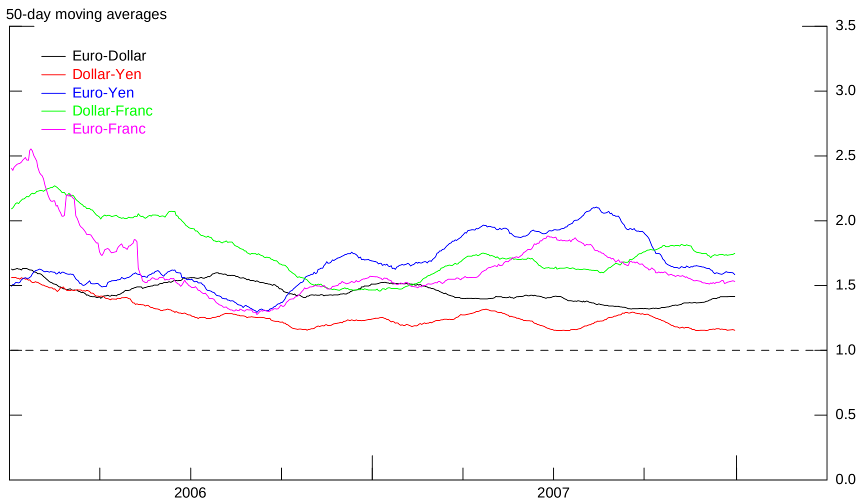

Figure 8: Relative Proportion of ‘Taking’ from Humans (Ratio of HC/CC to HH/CH)