

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

Alain Chaboud      Benjamin Chiquoine      Erik Hjalmarsson      Clara Vega\*

October 10, 2012

## Abstract

We study the impact of algorithmic trading in the foreign exchange market using a high-frequency dataset representing a majority of global interdealer trading in three major currency pairs, euro-dollar, dollar-yen, and euro-yen, from 2003 through 2007. We find that human-initiated trades account for a larger share of the variance in exchange rate returns than computer-initiated trades: humans are still the “informed” traders. There is some evidence, however, that algorithmic trading contributes to a more efficient price discovery process via the elimination of triangular arbitrage opportunities and the faster incorporation of macroeconomic news surprises into the price. We also show that algorithmic trades tend to be correlated, indicating that computer-driven strategies are not as diverse as those used by human traders. Despite this correlation, we find no evidence that algorithmic trading causes excess volatility. Furthermore, the amount of algorithmic activity in the market has a small, but positive, impact on market liquidity.

JEL Classification: F3, G12, G14, G15.

Keywords: Algorithmic trading; Liquidity provision; Price discovery; Private information.

---

\*Chaboud and Vega are with the Division of International Finance, Federal Reserve Board, Mail Stop 43, Washington, DC 20551, USA; Chiquoine is with the Investment Fund for Foundations, 97 Mount Auburn Street, Cambridge MA 02138, USA; Hjalmarsson is with Queen Mary, University of London, School of Economics and Finance, Mile End Road, London E1 4NS, UK. Please address comments to the authors via e-mail at [alain.p.chaboud@frb.gov](mailto:alain.p.chaboud@frb.gov), [bchiquoine@tiff.org](mailto:bchiquoine@tiff.org), [e.hjalmarsson@qmul.ac.uk](mailto:e.hjalmarsson@qmul.ac.uk), and [clara.vega@frb.gov](mailto:clara.vega@frb.gov). We are grateful to EBS/ICAP for providing the data, and to Nicholas Klagge and James S. Hebden for their excellent research assistance. We would like to thank Cam Harvey, an anonymous Associate Editor and an anonymous referee for their valuable comments. We also benefited from the comments of Gordon Bodnar, Charles Jones, Terrence Hendershott, Luis Marques, Albert Menkveld, Dagfinn Rime, Alec Schmidt, John Schoen, Noah Stoffman, and of participants in the University of Washington Finance Seminar, SEC Finance Seminar Series, Spring 2009 Market Microstructure NBER conference, San Francisco AEA 2009 meetings, the SAIS International Economics Seminar, the SITE 2009 conference at Stanford, the Barcelona EEA 2009 meetings, the Essex Business School Finance Seminar, the EDHEC Business School Finance Seminar, and the Imperial College Business School Finance Seminar. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

# 1 Introduction

The use of algorithmic trading, where computers monitor markets and manage the trading process at high frequency, has become common in major financial markets in recent years, beginning in the U.S. equity market in the late 1990s. Since the introduction of algorithmic trading, there has been widespread interest in understanding the potential impact it may have on market dynamics. While some have highlighted the potential for more efficient price discovery, others have expressed concern that it may lead to higher adverse selection costs and excessive volatility. Despite the widespread interest, formal empirical research on algorithmic trading has been rare, primarily because of a lack of data in which algorithmic trades are clearly identified.<sup>1</sup>

In this paper we analyze the effect algorithmic (“computer”) trades and non-algorithmic (“human”) trades have on the informational efficiency of foreign exchange prices; it is the first formal empirical study on the subject in the foreign exchange market. We rely on a novel dataset consisting of several years (October 2003 to December 2007) of minute-by-minute trading data from Electronic Broking Services (EBS) in three currency pairs: the euro-dollar, dollar-yen, and euro-yen. The data represent a majority of spot interdealer transactions across the globe in these exchange rates. A crucial feature of the data is that, on a minute-by-minute frequency, the volume and direction of human and computer trades are explicitly identified, allowing us to measure their respective impacts at high frequency. Another useful feature of the data is that it spans the introduction and rapid growth of algorithmic trading in a market where it had not been previously allowed.

The literature highlights two key features: algorithmic trading’s advantage in speed over human trading and the potentially high correlation in algorithmic traders’ strategies and actions. There is no agreement on the effect these two features of algorithmic trading may have on the informativeness of prices. Biais, Foucault, and Moinas (2011), and Martinez and Rosu (2011), show that algorithmic traders’ speed advantage over humans – their ability to react more quickly to public information than humans – has a positive effect on the informativeness of prices.<sup>2</sup> In their theoretical models algorithmic traders are better informed than humans and use market orders to exploit their information. Given these assumptions, the authors show that the presence of algorithmic traders makes asset prices more informationally efficient, but their trades are a source of adverse selection for those who provide liquidity. These authors argue that algorithmic traders

---

<sup>1</sup>A notable early exception prior to our study is the paper by Hendershott, Jones, and Menkveld (2011), who got around the data constraint by using the flow of electronic messages on the NYSE as a proxy for algorithmic trading. Subsequent to the first version of our study, several papers, e.g. Brogaard (2010), Hendershott and Riordan (2009, 2011), Hasbrouck and Saar (2010), Jovanovic and Menkveld (2011), Menkveld (2011), Kirilenko, Kyle, Samadi, and Tuzun (2010), and Zhang (2010), have conducted empirical work on the subject using stock market data. Most of these studies do not directly observe algorithmic or high frequency trading activity, but they infer the activity. Biais and Woolley (2011) and Foucault (2011) provide an excellent survey of the literature on algorithmic and high frequency trading.

<sup>2</sup>Public information is prices, quotes, and depth posted for that asset and other assets, news announced via Bloomberg or Reuters etc.

contribute to price discovery because once price inefficiencies exist, they quickly make them disappear. One can also argue that better informed algorithmic traders who specialize in providing liquidity make prices more informationally efficient by posting quotes that reflect new information quickly and thus prevent arbitrage opportunities from occurring in the first place (e.g., Hoffmann (2012)).

In contrast to these positive views on algorithmic trading, Foucault, Hombert, and Rosu (2012) show that in a world with no asymmetric information, algorithmic traders' speed advantage does not increase the informativeness of prices and simply increases adverse selection costs. Jarrow and Protter (2011) argue that both features of algorithmic trading, the speed advantage over human traders and the potential commonality of trading strategies amongst algorithmic traders, may have a negative effect on the informativeness of prices. In their theoretical model, algorithmic traders, triggered by a common signal, do the same trade at the same time. Algorithmic traders collectively act as one big trader, create price momentum and thus cause prices to be less informationally efficient. Kozhan and Wah Tham (2012) show that algorithmic traders entering the same trade at the same time causes a crowding effect, which in turn pushes prices further away from fundamentals. Stein (2009) also highlights this crowding effect in the context of hedge funds simultaneously implementing "convergence trade" strategies. In contrast, Oehmke (2009) and Kondor (2009) argue that the higher the number of traders who implement "convergence trade" strategies the more efficient prices will be. Foucault (2011), in his review of the literature, emphasizes the disagreement in the literature and concludes that the effect algorithmic traders have on the informativeness of prices ultimately depends on what are the trading strategies algorithmic traders specialize on and their investment in monitoring technologies. We contribute to this literature by estimating algorithmic traders' degree of correlated trading activity in the foreign exchange market and their role in the process of determining prices.

First we investigate whether the trading strategies used by computers are more correlated than those used by humans. Since we do not have specific data on trading strategies used by market participants at any point in time, we indirectly infer the correlation among computer trading strategies at a point in time from our trading data. The primary idea behind the correlation test that we design is that traders who follow similar trading strategies will trade less with each other than those who follow less correlated strategies. In a simple random matching model, the amount of trading between two types of traders is determined by the relative proportion of each type of trader in the market. Comparing the model's predictions to the realized values in the data, we find very strong evidence that computers do not trade with each other as much as the model would predict. There is thus evidence that the strategies embodied in the computer algorithms are indeed more correlated and less diverse than those used by human traders.

We next investigate the effect both algorithmic trades and the degree of correlation amongst algorithmic trading strategies have on a particular example of prices not being informationally efficient: the occurrence

of triangular arbitrage opportunities. We document that the introduction and growth of algorithmic trading coincides with a dramatic reduction in triangular arbitrage opportunities. We continue with a formal analysis of whether an increased (share of) algorithmic trading actually caused the reduction in triangular arbitrage opportunities, or whether the relationship is merely coincidental or due to the increase in market liquidity (trading volume) or the decrease in price volatility over time. To that end, we formulate a high-frequency vector autoregression (VAR) specification of the degree of triangular arbitrage opportunities and the degree of algorithmic trading activity and the degree of correlation amongst algorithmic trading strategies, controlling for trading volume and exchange rate return volatility in each currency pair. We estimate both a reduced form of the VAR and a structural VAR, which uses the heteroskedasticity identification approach developed by Rigobon (2003) and Rigobon and Sack (2003,2004).<sup>3</sup> In contrast to the reduced-form Granger causality tests, which essentially measure predictive relationships, the structural VAR estimation allows for an identification of the contemporaneous causal impact of algorithmic trading on triangular arbitrage opportunities.

Both the reduced form and structural VAR estimations show that algorithmic trading activity reduces the number of triangular arbitrage opportunities. In particular, we find that algorithmic traders predominantly reduce arbitrage opportunities by quickly acting on the posted quotes by humans that enable the profit opportunity. This result is consistent with the view that algorithmic trading improves informational efficiency by increasing the speed of price discovery, but at the same time they increase the adverse selection costs to slow traders as suggested by the theoretical models of Biais, Foucault, and Moinas (2011), and Martinez and Rosu (2011). Consistent with this result, we find that a higher degree of correlation amongst algorithmic trading strategies (or when computers are predominantly trading with humans more so than with other computers) reduces the number of arbitrage opportunities. Thus contrary to Jarrow and Protter (2011)'s model, in this particular example, commonality in trading strategies is helping the price discovery process. Interestingly, we also find some evidence that an increase in computers posting quotes decreases the number of triangular arbitrage opportunities. In other words, the mechanism described in the aforementioned papers is not the only way algorithmic traders are making prices more efficient, but there is some evidence that algorithmic traders are also making prices more efficient by posting quotes that reflect new information more quickly.

As explained above, we find that algorithmic traders reduce the number of arbitrage opportunities. However, this is just one example of how computers improve the price discovery process. It is possible that algorithmic traders reduce the number of arbitrage opportunities, but if during times when there are no triangular arbitrage opportunities algorithmic traders behave en masse like the positive-feedback traders of

---

<sup>3</sup>To identify the parameters of the structural VAR we use the heteroskedasticity in algorithmic trading activity across the sample.

DeLong, Shleifer, Summers, and Waldman (1990) or the chartists described in Froot, Scharfstein, and Stein (1992), or the short-term investors in Vives (1995), then they might cause deviations of prices from fundamental values and thus induce excess volatility. We thus investigate the effect algorithmic traders have on a more general measure of prices not being informationally efficiency: the magnitude of serial autocorrelation in currency returns.<sup>4</sup> In particular, we estimate the serial autocorrelation of 5-second returns over 5-minute intervals.<sup>5</sup> Similar to the evolution of arbitrage opportunities in the market, we document that the introduction and growth of algorithmic trading coincides with a reduction in the absolute value of serial autocorrelation in each of the three currency-pairs we analyze, especially in the euro-yen currency pair, the currency with the lowest trading volume. We estimate both a reduced form and a structural VAR and find that, on average, algorithmic trading participation reduces the degree of serial autocorrelation in currency returns. Interestingly, the improvement in the informational efficiency of prices comes from an increase in algorithmic traders provision of liquidity, not from an increase in algorithmic traders reaction to posted quotes. In other words, contrary to the previous example of triangular arbitrage opportunities, algorithmic traders appear to increase the informational efficiency of prices by posting quotes that reflect new information more quickly. Finally, we find that a higher correlation of algorithmic traders' actions is associated with an increase in the serial autocorrelation of currency returns, providing some support for Jarrow and Protter (2011)'s concern, namely the commonality in trading strategies may hinder the price discovery process, however the effect is not statistically significant.

The paper proceeds as follows. Section 2 introduces the high-frequency data used in this study, including a short description of the structure of the market and an overview of the growth of algorithmic trading in the foreign exchange market over time. Section 3 analyzes whether computer strategies are more correlated than human strategies. In Section ??, we test whether there is evidence that the share of algorithmic trading in the market has a causal impact on the informativeness of prices. Finally, Section 6 concludes. Some additional clarifying and technical material is found in the Appendix.

---

<sup>4</sup>Samuelson (1965), Fama (1965) and Fama (1970), among others, show that if prices reflect all public information, then they must follow a martingale process. As a consequence, an informationally efficient price exhibits no serial autocorrelation either positive (momentum) or negative (mean-reversion).

<sup>5</sup>Our choice of a 5-second frequency is driven by a trade-off between sampling at a high enough frequency that we estimate the effect algorithmic traders have on prices, but low enough that we have enough transactions to avoid a zero serial autocorrelation bias due to the number of zero-returns. Our results are qualitatively similar when we sample prices at 1-second for the most liquid currency pairs, euro-dollar and dollar-yen, but for the euro-yen the 1-second sampling frequency biases the serial auto-correlation towards zero due to a lack of trading activity at the beginning of the sample.

## 2 Data description

### 2.1 Market structure

Over our sample period, from 2003 to 2007, two electronic platforms processed the majority of global inter-dealer spot trading in the major currency pairs, one offered by Reuters, and one offered by Electronic Broking Services (EBS).<sup>6</sup> Both of these trading platforms are electronic limit order books. Importantly, trading in each major currency pair is highly concentrated on only one of the two systems. Of the most traded currency pairs (exchange rates), the top two, euro-dollar and dollar-yen, trade primarily on EBS, while the third, sterling-dollar, trades primarily on Reuters. As a result, price discovery for spot euro-dollar, for instance, occurs on the EBS system, and dealers across the globe base their customer and derivative quotes on that price. EBS controls the network and each of the terminals on which the trading is conducted. Traders can enter trading instructions manually, using an EBS keyboard, or, upon approval by EBS, via a computer directly interfacing with the system. The type of trader (human or computer) behind each trading instruction is recorded by EBS, allowing for our study.

The EBS system is an interdealer system accessible to foreign exchange dealing banks and, under the auspices of dealing banks (via prime brokerage arrangements), to hedge funds and commodity trading advisors (CTAs). As it is a “wholesale” trading system, the minimum trade size is 1 million of the “base” currency, and trade sizes are only allowed in multiple of millions of the base currency. We analyze data in the three most-traded currency pairs on EBS, euro-dollar, dollar-yen, and euro-yen.<sup>7</sup>

### 2.2 Price, volume, and order flow data

Our data consists of both quote data and transactions data. The quote data, at the one-second frequency, consist of the highest bid quote and the lowest ask quote on the EBS system in our three currency pairs. The quote data are available from 1997 through 2007. All the quotes are executable and therefore truly represent the market price at that instant.<sup>8</sup> From these data, we construct mid-quote series from which we can compute exchange rate returns at various frequencies. The transactions data, available from October 2003 through December 2007, are aggregated by EBS at the one-minute frequency. They provide detailed information on

---

<sup>6</sup>EBS, which was previously owned by a group of foreign-exchange dealing banks, was purchased by the ICAP group in 2006. ICAP owns interdealer trading platforms and voice-broking services for most types of financial assets in a number of countries.

<sup>7</sup>The euro-dollar currency pair is quoted as an exchange rate in dollars per euro, with the euro the “base” currency. Similarly, the euro is also the base currency for euro-yen, while the dollar is the base currency for the dollar-yen pair.

<sup>8</sup>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. Trading is continuous outside of the weekend, but the value date for trades, by convention, changes at 17:00 New York time, which therefore marks the end of each trading day. We also drop certain holidays and days of unusually light volume: December 24-December 26, December 31-January 2, Good Friday, Easter Monday, Memorial Day, Labor Day, Thanksgiving and the following day, and July 4 (or, if this is on a weekend, the day on which the U.S. Independence Day holiday is observed).

the volume and direction of trades that can be attributed to computers and humans in each currency pair. Specifically, each minute we observe trading volume and order flow for each of the four possible pairs of human and computer makers and takers: human-maker/human-taker ( $HH$ ), computer-maker/human-taker ( $CH$ ), human-maker/computer-taker ( $HC$ ), and computer-maker/computer-taker ( $CC$ ).<sup>9</sup>

Figure 1 shows, from 2003 through 2007, for each currency pair, the fraction of trading volume where at least one of the two counterparties is an algorithmic trader, i.e.,  $Vol(CH + HC + CC)$  as a fraction of total volume. From its beginning in late 2003, the fraction of trading volume involving algorithmic trading grows by the end of 2007 to near 60 percent for euro-dollar and dollar-yen trading, and to about 80 percent for euro-yen trading.

Figure 2 shows the evolution over time of the four different possible types of trades:  $Vol(HH)$ ,  $Vol(CH)$ ,  $Vol(HC)$ , and  $Vol(CC)$ , as fractions of the total volume. By the end of 2007, in the euro-dollar and dollar-yen markets, human to human trades, the solid lines, account for slightly less than half of the volume, and computer to computer trades, the dotted lines, for about ten to fifteen percent. In these two currency pairs,  $Vol(CH)$  is often slightly higher than  $Vol(HC)$ , i.e., computers “take” prices posted by humans, the dashed lines, less often than humans take prices posted by market-making computers, the dotted-dashed lines. The story is different for the cross-rate, the euro-yen currency pair. By the end of 2007, there are more computer to computer trades than human to human trades. But the most common type of trade in euro-yen is computers trading on prices posted by humans. We believe this reflects computers taking advantage of short-lived triangular arbitrage opportunities, where prices set in the euro-dollar and dollar-yen markets, the primary sites of price discovery, are very briefly out of line with the euro-yen cross rate. Detecting and trading on triangular arbitrage opportunities is widely thought to have been one of the first strategies implemented by algorithmic traders in the foreign exchange market, which is consistent with the more rapid growth in algorithmic activity in the euro-yen market documented in Figure 1. We discuss the evolution of the frequency of triangular arbitrage opportunities in Section 4 below.

### 3 How Correlated Are Algorithmic Trades and Strategies?

Jarrow and Protter (2011) argue that the potential commonality of trading strategies amongst algorithmic traders may have a negative effect on the informativeness of prices. In their theoretical model, algorithmic traders, triggered by a common signal, do the same trade at the same time. Algorithmic traders collectively act as one big trader, create price momentum and thus cause prices to be less informationally efficient.

---

<sup>9</sup>The naming convention for “maker” and “taker” reflects the fact that the “maker” posts quotes before the “taker” chooses to trade at that price. Posting quotes is, of course, the traditional role of the market-“maker.” We refer the reader to Appendix A1 for more details on how we calculate volume and order flow for these four possible pairs of human and computer makers and takers. Order flow is defined as the net of buyer-initiated trading volume minus seller-initiated trading volume.

Similarly, Khandani and Lo (2007, 2011), who analyze the large losses that occurred for many quantitative long-short equity strategies at the beginning of August 2007, highlight the possible adverse effects on the market of such commonality in behavior across market participants (algorithmic or not) and provide empirical support for this concern.

If one looks for similar episodes in our data, August 16, 2007 in the dollar-yen market stands out. It is the day with the highest realized volatility and one of the highest absolute value of serial correlation in 5-second returns 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 shown in Figure 3. The figure also shows, for each 30-minute interval in the day, computer-taker order flow ( $HC + CC$ ) in the top panel and human-taker order flow ( $HH + CH$ ) in the lower panel. The two sharp exchange rate movements mentioned happened when computers, as a group, aggressively initiated sales of dollars and purchases of yen. Computers, during these periods of sharp yen appreciation, mainly traded with humans, not with other computers. Human order flow at those times was, in contrast, quite small, even though the trading volume initiated by humans (not shown) was well above that initiated by computers: human takers were therefore selling and buying dollars in almost equal amounts. The orders initiated by computers during those time intervals were therefore far more correlated than the orders initiated by humans. After 12:00 p.m., human traders, in aggregate, began to buy dollars fairly aggressively, and the appreciation of the yen against the dollar was partially reversed.

The August 16, 2007 episode in the dollar-yen market was widely viewed at the time as the result of a sudden unwinding of large yen carry-trade positions, with many hedge funds and banks' proprietary trading desks closing risky positions and buying yen to pay back low-interest loans.<sup>10</sup> This is, of course, only one episode in our two-year sample, and by far the most extreme as to its impact on volatility, so one should not draw conclusions about the overall correlation of algorithmic strategies based on this single instance. Furthermore, episodes of very sharp appreciation of the yen due to the rapid unwinding of yen carry trades have occurred on a number of occasions since the late 1990s, well before algorithmic trading was allowed in this market.<sup>11</sup> We therefore investigate next whether there is evidence that, on average over our sample, the strategies used by computer traders have tended to be more correlated (less diverse) than those used by human traders.<sup>12</sup>

---

<sup>10</sup>A traditional carry-trade strategy borrows in a low-interest rate currency and invests in a high-interest rate currency, with the implicit assumption that the interest rate differential will not be (fully) offset by changes in the exchange rate. That is, carry trades bet on uncovered interest rate parity not holding. Although the August 16, 2007 episode occurs only a week after the events described in Khandani and Lo (2007, 2011), we are not aware of any direct link between the quant equity crisis and the carry trade unwinding.

<sup>11</sup>The sharp move of the yen in October 1998, which included a 1-day appreciation of the yen against the dollar of more than 7 percent, is the best-known example of the impact of the rapid unwinding of carry trades.

<sup>12</sup>There is little public knowledge, and no data, about the mix of strategies used by algorithmic traders in the foreign exchange market, as traders and EBS keep what they know confidential. From conversations with market participants, we believe that about half of the algorithmic trading volume on EBS over our sample period comes from traditional foreign exchange dealing banks, with the other half coming from hedge funds and commodity trading advisors (CTAs). Hedge funds and CTAs, who access EBS under prime-brokerage arrangements, can only trade algorithmically (no keyboard trading) over our sample period. Some of



### 3.1 Inferring Correlation from Trade Data

We do not observe the trading strategies of market participants. However, we can infer the correlation of algorithmic strategies from the trading activity of computers and humans. The idea is the following. Traders who follow similar trading strategies and therefore send similar trading instructions at the same time, will trade less with each other than those who follow less correlated strategies. Therefore, the extent to which computers trade with each other contains information about how correlated the algorithmic strategies are.

More precisely, we consider a simple benchmark model that assumes random and independent matching of traders. This is a reasonable assumption given the lack of discrimination between keyboard traders and algorithmic traders in the EBS matching process; that is, EBS does not differentiate in any way between humans and computers when matching buy and sell orders in its electronic order book. The model allows us to determine the theoretical probabilities of the four possible trades: Human-maker/human-taker, computer-maker/human-taker, human-maker/computer-taker and computer-maker/computer-taker. We then compare these theoretical probabilities to those observed in the actual data. The benchmark model is fully described in Appendix A3, and below we outline the main concepts and empirical results.

Under our random and independent matching assumption, computers and humans, both of which are indifferent *ex-ante* between making and taking, trade with each other in proportion to their relative presence in the market. In a world with more human trading activity than computer trading activity (which is the case in our sample), we should observe that computers take more liquidity from humans than from other computers. That is, the probability of observing human-maker/computer-taker trades,  $Prob(HC)$ , should be larger than the probability of observing computer-maker/computer taker trades,  $Prob(CC)$ . We label the ratio of the two,  $Prob(HC)/Prob(CC)$ , the computer-taker ratio,  $RC$ . Similarly, one expects humans to take more liquidity from other humans than from computers, i.e.,  $Prob(HH)$  should be larger than  $Prob(CH)$ . We label this ratio,  $Prob(HH)/Prob(CH)$ , the human-taker ratio,  $RH$ . In summary, one thus expects that  $RC > 1$  and  $RH > 1$ , because there is more human trading activity than computer trading activity.

Importantly, the model predicts that the ratio of these two ratios, the computer-taker ratio divided by the human-taker ratio, should be equal to one. That is, the model predicts  $R = RC/RH = 1$  because humans take liquidity from other humans in the same *proportion* that computers take liquidity from humans.

Observing a ratio  $R = RC/RH > 1$  in the data indicates that computers are trading less among themselves

---

the banks' computer trading is related to activity on their own customer-to-dealer platforms, to automated hedging activity, and to the optimal execution of large orders. But a sizable fraction (perhaps almost a half) is believed to be proprietary trading using a mix of strategies similar to what hedge funds and CTAs use. These strategies include various types of high-frequency arbitrage, including across different asset markets, a number of lower-frequency statistical arbitrage strategies (including carry trades), and strategies designed to automatically react to news and data releases (believed to be still fairly rare by 2007). Overall, market participants believe that the main difference between the mix of algorithmic strategies used in the foreign exchange market and the mix used in the equity market is that optimal execution algorithms are less prevalent in foreign exchange than in equity.

and more with humans than what our benchmark model predicts.<sup>13</sup> Therefore a ratio bigger than one can be viewed as evidence that computers have trading strategies that are more correlated than those of humans.

The test outlined above implicitly takes into account trading direction because the matching process in EBS takes it into account. Nevertheless, we also describe in detail in Appendix A3 a model that explicitly takes trading direction into account. Using notation similar to the model without trading direction, this model yields four ratios,  $RC^B$ ,  $RC^S$ ,  $RH^B$ , and  $RH^S$ , a computer-taker ratio where computers are buying, a computer-taker ratio where computers are selling, a human-taker ratio where humans are buying, and a human-taker ratio where humans are selling. As before, the model predicts that each of these four ratios will be greater than one, but that the ratio of the buy ratios,  $R^B \equiv \frac{RC^B}{RH^B}$ , and the ratio of the sell ratios,  $R^S \equiv \frac{RC^S}{RH^S}$ , will both be equal to one.

Based on the model described above, we calculate for each trading day in our sample a realized value for daily, 1-min, and 5-min  $R$ ,  $R^S$ , and  $R^B$ . Specifically, the daily realized values of  $RH$  and  $RC$  are given by  $\widehat{RH} = \frac{Vol(HH)}{Vol(CH)}$  and  $\widehat{RC} = \frac{Vol(HC)}{Vol(CC)}$ , where, for instance,  $Vol(HC)$  is the daily trading volume between human makers and computer takers, following the notation described in Section 2. Similarly, we define  $\widehat{RH}^S = \frac{Vol(HH^S)}{Vol(CH^S)}$ ,  $\widehat{RH}^B = \frac{Vol(HH^B)}{Vol(CH^B)}$ ,  $\widehat{RC}^S = \frac{Vol(HC^S)}{Vol(CC^S)}$ , and  $\widehat{RC}^B = \frac{Vol(HC^B)}{Vol(CC^B)}$ , where  $Vol(HH^B)$  is the daily buy volume between human makers and human takers (i.e., buying of the base currency by the taker),  $Vol(HH^S)$  is the daily sell volume between human makers and human takers, and so forth.

Table 3 shows the means of the natural log of 1-min, 5-min, and daily ratios of ratios,  $\ln \widehat{R} = \ln(\frac{\widehat{RC}}{\widehat{RH}})$ ,  $\ln \widehat{R}^S = \ln(\frac{\widehat{RC}^S}{\widehat{RH}^S})$ , and  $\ln \widehat{R}^B = \ln(\frac{\widehat{RC}^B}{\widehat{RH}^B})$ , for each currency pair. In contrast to the benchmark predictions that  $R \equiv 1$ ,  $R^B \equiv 1$  and  $R^S \equiv 1$ , or equivalently that  $\ln R \equiv 0$ ,  $\ln R^B \equiv 0$  and  $\ln R^S \equiv 0$ , we find that, for all three currency pairs, at all frequencies,  $\ln \widehat{R}$ ,  $\ln \widehat{R}^B$  and  $\ln \widehat{R}^S$  are substantially and significantly greater than zero.<sup>14</sup> The table also shows the number of intervals in which the statistics are above zero. In all currencies, at the daily frequency, more than 95 percent of the days,  $\ln \widehat{R}$ ,  $\ln \widehat{R}^B$  and  $\ln \widehat{R}^S$  are above zero. Overall, the observed ratios are highest in the cross-rate, the euro-yen, consistent with the view that computers trading in the cross-rate are predominantly taking advantage of short-lived triangular arbitrage opportunities, and thus are more likely to do the same trade at the same time in this currency pair.

We also show in Table 3 the number of non-missing observations for the ratio of the ratios. At high frequencies, 1-minute, 5-minute, the majority of the observations are missing because it is common for

<sup>13</sup>For the ratio  $R$  to be larger than one, either the computer-taker ratio is larger than what the model predicts or the human-taker ratio is smaller than the model predicts. For the computer-taker ratio to be larger than what the model predicts, either computers are taking too much liquidity from humans, or computers are not taking enough liquidity from other computers. Similarly, if the human-taker ratio is smaller than what the model predicts, computers are trading more with humans than the model predicts.

<sup>14</sup>We report summary statistics for  $\ln \widehat{R}$ ,  $\ln \widehat{R}^S$  and  $\ln \widehat{R}^B$  rather than  $\widehat{R}$ ,  $\widehat{R}^S$  and  $\widehat{R}^B$  because  $\widehat{R}$ ,  $\widehat{R}^S$  and  $\widehat{R}^B$  are bounded below by zero, while  $\ln \widehat{R}$ ,  $\ln \widehat{R}^S$  and  $\ln \widehat{R}^B$  are not bounded. Furthermore, the Taylor expansion of  $\ln \widehat{R}$ ,  $\ln \widehat{R}^S$  and  $\ln \widehat{R}^B$  is simpler. However, our results are robust to using  $\widehat{R}$ ,  $\widehat{R}^S$  and  $\widehat{R}^B$  or  $\ln \widehat{R}$ ,  $\ln \widehat{R}^S$  and  $\ln \widehat{R}^B$ .

$Vol(CC)$ ,  $Vol(CC^S)$ , and  $Vol(CC^B)$  to be zero at high frequencies, especially at the beginning of our sample period when algorithmic trading was scarce.<sup>15</sup> The fact that we cannot estimate  $\ln \widehat{R}$ ,  $\ln \widehat{R^B}$  and  $\ln \widehat{R^S}$  when  $Vol(CC)$ ,  $Vol(CC^S)$ , and  $Vol(CC^B)$  are zero, introduces a downward bias in our sample statistics at high frequencies because  $\ln \widehat{R}$ ,  $\ln \widehat{R^B}$  and  $\ln \widehat{R^S}$  are large when  $Vol(CC)$  is close to zero. In other words,  $\ln \widehat{R}$ ,  $\ln \widehat{R^B}$  and  $\ln \widehat{R^S}$  underestimates how highly correlated algorithmic trading strategies are at the 1-minute and 5-minute frequency. The daily  $\ln \widehat{R}$ ,  $\ln \widehat{R^B}$  and  $\ln \widehat{R^S}$  values, which we are able to compute on 80 percent of the days, do not suffer from this bias as much and accordingly show that algorithmic trading strategies are more highly correlated (i.e.,  $\ln \widehat{R}$ ,  $\ln \widehat{R^B}$  and  $\ln \widehat{R^S}$  are farther from zero at the daily frequency than at the 1-minute and 5-minute frequency).

Another concern, is that we observe  $\ln \widehat{R}$  during periods of high trading volume — when none of the following quantities:  $Vol(CC)$  nor  $Vol(HH)$  nor  $Vol(CH)$  are equal to zero. To mitigate these concerns we show in Table A1 in the appendix the means of  $\ln \widehat{R}$ ,  $\ln \widehat{R^S}$  and  $\ln \widehat{R^B}$  at all frequencies using data from 2006 to 2007, when the number of missing observations for each of these ratios is smaller. The test statistics show that the means of  $\ln \widehat{R}$ ,  $\ln \widehat{R^S}$  and  $\ln \widehat{R^B}$  are robustly above zero. Importantly, the percent of days  $\ln \widehat{R}$ ,  $\ln \widehat{R^S}$  and  $\ln \widehat{R^B}$  are above zero is from 98 to 100 percent, suggesting that algorithmic trading activity is highly correlated during high, medium and low trading volume days. To further mitigate concerns we compute first order Taylor expansion approximations of  $\ln \widehat{R}$ ,  $\ln \widehat{R^S}$  and  $\ln \widehat{R^B}$  around mean values of  $Vol(CC)$ ,  $Vol(HH)$ ,  $Vol(HC)$  and  $Vol(CH)$ . This approximation will be particularly useful at the 1-minute and 5-minute frequency. Specifically, each day we compute the mean of  $Vol(CC)$ ,  $Vol(HH)$ ,  $Vol(HC)$  and  $Vol(CH)$  at the 1-minute and 5-minute frequency. Each month we compute the mean of  $Vol(CC)$ ,  $Vol(HH)$ ,  $Vol(HC)$  and  $Vol(CH)$  at the daily frequency. We then compute  $\ln \widehat{R}$  at each frequency using a first order Taylor expansion approximation. The first order Taylor expansion of  $\ln \widehat{R} = \ln(Vol(HC)) - \ln(Vol(CC)) - \ln(Vol(HH)) + \ln(Vol(CH))$  around the values  $a = \overline{Vol(HC)}$ ,  $b = \overline{Vol(CC)}$ ,  $c = \overline{Vol(HH)}$ , and  $d = \overline{Vol(CH)}$ , is  $\ln \widehat{R} \approx \ln(a) + \frac{1}{a}(Vol(HC) - a) - \frac{1}{b}(Vol(CC) - b) - \frac{1}{c}(Vol(HH) - c) + \frac{1}{d}(Vol(CH) - d)$

We report in Table 3 the mean of  $\ln \widehat{R}$ ,  $\ln \widehat{R^S}$  and  $\ln \widehat{R^B}$  when we replace missing values  $\ln \widehat{R}$ ,  $\ln \widehat{R^S}$  and  $\ln \widehat{R^B}$  with the first order Taylor expansion approximations of  $\ln \widehat{R}$ ,  $\ln \widehat{R^S}$  and  $\ln \widehat{R^B}$  described above at each frequency. When we do this we are able to observe  $\ln \widehat{R}$ ,  $\ln \widehat{R^S}$  and  $\ln \widehat{R^B}$  for more than 80 percent of the observations at each frequency in our full sample and for 100 percent of the observations in the 2006-2007 sample (results shown in Table B1 in the Appendix). The results shown in Table 3, Table A1 and Table

<sup>15</sup> $\ln R$  is missing if either  $Vol(CC)$  or  $Vol(HH)$  or  $Vol(CH)$  is equal to zero. However, the probability that  $Vol(CC)$  is equal to zero is higher than the probability that either  $Vol(HH)$  or  $Vol(CH)$  are equal to zero. In our full sample,  $Vol(CC)$  is equal to zero in 70 percent, 74 percent, and 76 percent of our 1-minute observations in EUR/USD, USD/JPY and EUR/JPY, respectively. In contrast,  $Vol(HH)$  is equal to zero less than one percent of the time in the EUR/USD currency pair, 5 percent of the time in the USD/JPY currency pair, and 25 percent of the time in the EUR/JPY currency pair.  $Vol(CH)$  is equal to zero 30 percent and 40 percent of the time in the EUR/USD and USD/JPY currency pairs, respectively and 60 percent of the time in the EUR/JPY currency pair.

B1 suggest that despite our concern with missing observations, our conclusions are qualitatively the same. For all currencies at all frequencies the means of  $\ln \widehat{R}$ ,  $\ln \widehat{R}^S$  and  $\ln \widehat{R}^B$  are robustly above zero. The Tables also confirm the downward bias of our sample statistics estimated using the full sample and without using a Taylor approximation. The means of  $\ln \widehat{R}$ ,  $\ln \widehat{R}^S$  and  $\ln \widehat{R}^B$  are higher in the 2006 to 2007 sample, when  $Vol(CC)$  is less often zero, than in the full sample, September 2003 to December 2007. The means of  $\ln \widehat{R}$ ,  $\ln \widehat{R}^S$  and  $\ln \widehat{R}^B$  are higher when we use Taylor approximations, when  $\ln \widehat{R}$ ,  $\ln \widehat{R}^S$  and  $\ln \widehat{R}^B$  are missing less often, than when we do not use the Taylor approximations.<sup>16</sup>

In summary, the results show that computers do not trade with each other as much as random matching would predict, and the results hold for high, medium and low trading volume intervals. We take this as evidence that the algorithmic trading strategies used by computers are less diverse than the trading strategies used by human traders. Although high correlations among computer strategies may raise some concerns about the impact of algorithmic trading on the foreign exchange market, a high correlation of algorithmic strategies need not necessarily be detrimental to market quality. For instance, as noted above, the evidence for a high correlation of algorithmic strategies is strongest for the euro-yen currency pair. This is consistent with a large fraction of algorithmic strategies in that currency pair being used to detect and exploit triangular arbitrage opportunities. Faced with the same price data at a particular moment, the various computers seeking to profit from the same arbitrage opportunities would certainly take the same side of the market. However, this can contribute to a more efficient price discovery process in the euro-yen market, as suggested by Oehmke (2009) and Kondor (2009), or it can have adverse effects on the price discovery process, as suggested by Jarrow and Protter (2011), Kozhan and Wah Tham (2012) and Stein (2009). If the high correlation of strategies reflects a large number of algorithmic traders using the same carry trade or momentum strategies, as in the August 2007 example shown at the beginning of this section, then there may be reasons for concern. In the next sections we analyze the potential effects of algorithmic trading participation and a high correlation among AT strategies on triangular arbitrage opportunities and the serial correlation of high frequency returns.

## 4 Triangular Arbitrage and AT

The most distinguishing feature of algorithmic trading is the speed at which it can operate. Computers can both execute a given trade order as well as process relevant information at a much quicker pace than a human trader. Increased algorithmic trading might therefore help improve the speed with which information is incorporated into prices, as suggested by Biais, Foucault, and Moinas (2011), Martinez and Rosu (2011), Oehmke (2009) and Kondor (2009). However, algorithmic traders, triggered by a common signal, doing the

---

<sup>16</sup>We also conducted the same tests using statistics based on the number of trades of each type (HC, for instance) rather than trading volume of each type. The results were qualitatively identical.

same trade at the same time may hinder price discovery, as suggested by Jarrow and Protter (2011), Kozhan and Wah Tham (2012) and Stein (2009). We test this hypothesis by analyzing a particular situation where speed is of the essence: The capturing of triangular arbitrage opportunities. We begin with providing some suggestive graphical evidence that the introduction and growth of AT coincides with a reduction on triangular arbitrage opportunities, and then proceed with a more formal analysis.

## 4.1 Preliminary graphical evidence

Our data contains second-by-second bid and ask quotes on three related exchange rates (euro-dollar, dollar-yen, and euro-yen), and thus we can estimate the frequency with which these exchange rates are “out of alignment.” More precisely, each second we evaluate whether a trader, starting with a dollar position, could profit from purchasing euros with dollars, purchasing yen with euros, and purchasing dollars with yen, all simultaneously at the relevant bid and ask prices. An arbitrage opportunity is recorded for any instance when such a strategy (and/or a “round trip” in the other direction) would yield a profit of one basis point or more.<sup>17</sup> The daily frequency of such opportunities is shown from 2003 through 2007 in Figure 4.

The frequency of arbitrage opportunities drops dramatically over our sample, with the drop being particularly noticeable around 2005, when the rate of growth in algorithmic trading is highest. On average in 2003 and 2004, the frequency of such arbitrage opportunities is about 0.5 percent, one occurrence every 500 seconds. By 2007, at the end of our sample, the frequency has declined to 0.03 percent, one occurrence every 30,000 seconds. This simple analysis highlights the potentially important impact of algorithmic trading in this market. It is clear that other factors could have contributed to, or even driven, the drop in arbitrage opportunities, and the analysis certainly does not prove that algorithmic trading caused the decline. However, the findings line up well with the anecdotal (but widespread) evidence that one of the first strategies widely implemented by algorithmic traders in the foreign exchange market aimed to detect and profit from triangular arbitrage opportunities and the view that the more arbitrageurs there are taking advantage of the opportunities the less of these opportunities there are (Oehmke (2009) and Kondor (2009)).

---

<sup>17</sup>We conduct our test over the busiest period of the trading day in these exchange rates, from 3:00 am to 11:00 am New York Time, when all three exchange rates are very liquid. Our choice of a one-basis-point profit cutoff (\$100 per \$1 million traded) is arbitrary but, we believe, reasonable; the frequencies of arbitrage opportunities based on several other minimum profit levels (zero and 0.5 basis point) or higher profit levels (2 basis points) show a similar pattern of decline over time. Note that even though we account for actual bid and ask prices in our calculations of profits, an algorithmic trader also incurs other costs (e.g., overhead, fees for the EBS service, and settlement fees). In addition, the fact that trades on the system can only be made in whole millions of the base currency creates additional uncertainty and implied costs. As an example, if a trader sells 2 million dollars for 1.5 million euros, the next leg of the triangular arbitrage trade on EBS can only be a sale of, say, 1 or 2 million euros for yen, not a sale of 1.5 million euros. Therefore, even after accounting for bid-ask spreads, setting a minimum profit of zero to detect triangular arbitrage opportunities is not realistic.

## 4.2 A formal analysis

The evidence presented in Figure 4 is suggestive of the impact of algorithmic trading on informational efficiency. However, other factors might have influenced the speed with which information is incorporated into prices, and we now attempt to more formally identify the role of AT. Since triangular arbitrage is a high-frequency phenomenon, we perform the econometric analysis using minute-by-minute data, the highest frequency at which we sample algorithmic trading activity. The second-by-second quote data are used to construct a minute-by-minute measure of the frequency of triangular arbitrage opportunities. In particular, following the same approach as above, for each second we calculate the maximum profit achievable from a “round trip” triangular arbitrage trade in either direction, starting with a dollar position. This measure is truncated at zero, such that profits are always non-negative, and a minute-by-minute measure of triangular arbitrage opportunities is calculated as the number of seconds within each minute with a positive profit.

Algorithmic trading activity is measured on a minute-by-minute frequency in five different ways. First, the fraction of total volume of trade that involves a computer on at least one side of the trade, which we label  $VAT = 100 \times \frac{Vol(HC)+Vol(CH)+Vol(CC)}{Vol(HH)+Vol(HC)+Vol(CH)+Vol(CC)}$ . Second, the relative taking activity of computers, which we label  $VCT = 100 \times \frac{Vol(HC)+Vol(CC)}{Vol(HH)+Vol(HC)+Vol(CH)+Vol(CC)}$ . Third, the relative making activity of computers, which we label  $VCM = 100 \times \frac{Vol(CH)+Vol(CC)}{Vol(HH)+Vol(HC)+Vol(CH)+Vol(CC)}$ . Fourth, the relative taking activity of computers taking into account the sign of the trades, which we label  $OFCT = 100 \times \frac{|OF(C-Take)|}{|OF(C-Take)|+|OF(H-Take)|}$ . Fifth,  $\ln \widehat{R}$  as a measure of how highly correlated computer trading strategies or trading actions are. The higher the value of  $\ln \widehat{R}$  the higher the correlation of computer trading actions. As mentioned above,  $\ln \widehat{R}$  is missing for the majority of one-minute intervals. To mitigate concerns that arise from missing observations of  $\ln \widehat{R}$ , we also estimate our results using the Taylor expansion approximation of  $\ln \widehat{R}$  whenever we cannot observe  $\ln \widehat{R}$ . The results are qualitatively similar and available from the authors upon request.

Algorithmic trading and triangular arbitrage opportunities are likely determined simultaneously, in the sense that both variables have a contemporaneous impact on each other. OLS regressions of contemporaneous triangular arbitrage opportunities on contemporaneous algorithmic trading activity are therefore likely biased and misleading. In order to overcome these difficulties, we estimate a structural VAR system, which will be identified through the heteroskedasticity identification approach developed by Rigobon (2003) and Rigobon and Sack (2003, 2004). Let  $Arb_t$  be the minute-by-minute measure of triangular arbitrage possibilities and  $AT_t^{avg}$  be average AT activity across the three currency pairs;  $AT_t^{mean}$  is used to represent either of the five measures of AT activity. Define  $Y_t = (Arb_t, AT_t^{avg})$  and the structural form of the system is given by

$$AY_t = \Phi(L)Y_t + \Lambda X_{t-1:t-20} + \Psi G_t + \epsilon_t. \quad (1)$$

Here  $A$  is a  $4 \times 4$  matrix specifying the contemporaneous effects, normalized such that all diagonal elements are equal to 1.  $\Phi(L)$  is a lag-function that controls for the effects of the lagged endogenous variables.  $X_{t-1:t-20}$  consists of lagged control variables not modelled in the VAR. Specifically,  $X_{t-1:t-20}$  includes the sum of the volume of trade in each currency pair over the past 20 minutes and the volatility in each currency pair over the past 20 minutes, calculated as the sum of absolute returns over these 20 minutes.  $G_t$  represent a set of deterministic functions of time  $t$ , capturing individual trends and intra-daily patterns in the variables in  $Y_t$ . In particular,  $G_t = (I_{\{t \in 1\text{st month of sample}\}}, \dots, I_{\{t \in \text{last month of sample}\}}, I_{\{t \in 1\text{st half-hour of day}\}}, \dots, I_{\{t \in \text{last half-hour of day}\}})$ , capturing long term secular trends in the data by year-month dummy variables, as well as intra-daily patterns captured by half-hour dummy variables.<sup>18</sup> The structural shocks to the system are given by  $\epsilon_t$ , which, in line with standard structural VAR assumptions, are assumed to be independent of each other and serially uncorrelated at all leads and lags. The number of lags included in the VAR is set to 20. Equation (1) thus provides a very general specification, allowing for full contemporaneous interaction between the two variables, arbitrage opportunities and one of the five measures of AT activity.

Before describing the estimation of the structural system, it is useful to begin the analysis with the reduced form system,

$$Y_t = A^{-1}\Phi(L)Y_t + A^{-1}\Lambda X_{t-1:t-20} + A^{-1}\Psi G_t + A^{-1}\epsilon_t. \quad (2)$$

The reduced form is estimated equation-by-equation using ordinary least squares, and Granger causality tests are performed to assess the role of AT in determining triangular arbitrage opportunities. In particular, we test whether the *sum* of the coefficients on the lags of the causing variable is equal to zero. Since the sum of the coefficients on the lags of the causing variable is proportional to the long-run impact of that variable, the test can be viewed as a long-run Granger causality test. Importantly, the sum of the coefficients also tells us of the estimated direction of the (long-run) relationship, such that the test is associated with a clear direction in the causation.

Table ?? shows the results, where the first three rows in each sub-panel provides the Granger causality results, showing the sum of the coefficients on the lags of the causing variable, as well as the F-statistic and corresponding p-value.<sup>19</sup> The next two rows in each sub-panel provides the results of the standard Granger test, namely that all of the coefficients are equal to zero. The left hand panels show tests of whether AT causes (a reduction of) triangular arbitrage opportunities, whereas the right hand panels show test of whether triangular arbitrage opportunities have an impact on algorithmic trading. We estimate the VAR

<sup>18</sup>Replacing the year-month dummy variables with a linear and quadratic trend in  $t$ , yielded very similar estimation results.

<sup>19</sup>We also conducted standard Granger causality tests, which are simply F-tests of whether the coefficients on all lags of the causing variable are jointly equal to zero. Since this type of test is not associated with a clear direction in causation, we focus on the long-run test reported in Table ??, which explicitly shows the direction of causation. Overall, the traditional Granger causality tests yielded very similar results to the long-run test.

using two different samples: (i) using only data during the busiest trading hours of the day, between 3am and 11am New York time, and (ii) using data from the entire 24 hour trading day. We show the results only for the busiest trading hours of the day, between 3am and 11am New York time. The results using data from the entire 24 hour trading day are available from the authors upon request and are qualitatively similar. The (long-run) Granger causality tests of whether AT has a causal effect on triangular arbitrage all tell a fairly clear story. The sum of the coefficients on lagged AT activity is almost always negative and often statistically significant, indicating that increased AT activity leads to fewer triangular arbitrage opportunities. There is also evidence that an increase in triangular arbitrage opportunities Granger causes an increase in algorithmic trading. Importantly, the evidence is strongest for triangular arbitrage opportunities causing an increase in the relative taking activity of computers (*VCt* and *OFct* are strongly associated with a decrease in triangular arbitrage opportunities), which suggests that algorithmic traders improve the informational efficiency of prices by taking advantage of arbitrage opportunities and making them disappear quickly, rather than by posting quotes that prevent these opportunities from occurring (as would be indicated by a strong effect of *VCm* on reducing triangular arbitrage opportunities). This is in line with the theoretical models of Biais, Foucault, and Moinas (2011), Martinez and Rosu (2011), Oehmke (2009) and Kondor (2009).

The Granger causality results point to a potentially strong causal relationship between AT and triangular arbitrage, with causation seemingly going in both directions and in line with theory. However, although Granger causality tests can be quite informative and possibly quite indicative of true causality, they are based on the reduced form of the VAR, and do not explicitly identify the *contemporaneous* causal economic relationships in the model. This is particularly important in our setting, because we only observe trading activity at the 1-minute frequency. We therefore also attempt to estimate the structural version of the model, using a version of the heteroskedasticity identification approach developed by Rigobon (2003) and Rigobon and Sack (2003,2004). The basic idea of this identification scheme is that heterogeneity in the error terms can be used to identify simultaneous equation systems. The actual mechanics of the identification scheme are provided in the Appendix, and here we try to provide some intuition on how it works.

To form an idea of how the identification method works, consider, for simplicity, a system of two simultaneous equations, say triangular arbitrage opportunities and a measure of algorithmic trading in a single currency pair. In a typical classical setup, the impact of algorithmic trading on triangular arbitrage cannot be identified because of the contemporaneous feedback between the two variables. Now, suppose that algorithmic trading is more variable in the second half of the sample than in the first half, whereas the variance of triangular arbitrage opportunities remain the same. In this case, one can use the difference in variance in algorithmic trading across the two subsamples to identify the causal impact of algorithmic trading on triangular arbitrage. In particular, in the higher variance period the (causal) impact of algorithmic trading



on triangular arbitrage opportunities will be more important in determining the (“non-causal”) covariance between AT and triangular arbitrage, than in the lower variance period. That is, the reduced form covariance between AT and triangular arbitrage is a function of the variances of the structural shocks and the causal impact that each variable has on the other. A shift in the variance of AT therefore provides sufficient additional information to identify its causal impact on triangular arbitrage. More generally, the approach allows for a full identification of the simultaneous system, provided the covariance matrix changes in a non-proportional manner across two different variance regimes.

In the current context, we rely on the observation that the degree of AT participation in the market becomes more variable over time. That is, the variance of any of the measures of AT, such as the fraction of traded volume that involves a computer in some manner, tends to increase over time as this fraction becomes larger; as the level increases, so does the variance. To capitalize on this fact, we split up the sample into two equal-sized sub-samples, simply defined as the first and second half of the sample period. Although the variances of the shocks are surely not constant within these two subsamples, this is not crucial for the identification mechanism to work, as long as there is a clear distinction in (average) variance across the two sub-samples, as discussed in detail in Rigobon (2003). Rather, the crucial identifying assumption is that the structural parameters determining the contemporaneous impact between the variables (i.e.,  $A$  in equation (1) above) is constant across the two variance regimes. This is, of course, a strong assumption, although it is almost always implicitly made in any model with constant coefficients across the entire sample. To the extent that one attempts to make use of a long data span to identify the effects of algorithmic trading, it would be very difficult to do so without making some implicit assumption that the underlying structural impact is similar across the sample period. The Appendix provides more detail on the identification approach and details the mechanics of the actual estimation, which is performed via GMM. The Appendix also lists estimates of the covariance matrices across the two different regimes, showing that there is strong heteroskedasticity between the first and second half of the sample, and that the shift in the covariance matrices between the two regimes is not proportional.

Estimates of the relevant contemporaneous structural parameters in equation (1) are shown at the bottom of each sub-panel in Table ??, with bootstrapped standard errors given in parentheses below (see Appendix for details on the bootstrap). Overall, the contemporaneous effects are in line with those found in the Granger causality tests, namely increased AT activity, especially an increase in the relative taking activity of computers, causes a reduction in triangular arbitrage opportunities. Interestingly, the structural estimation indicates that an increase in correlated trading actions by computers is most effective in decreasing triangular arbitrage opportunities, consistent with the view that the more arbitrageurs there are in the market the sooner the arbitrage opportunities disappear (Oehmke (2009) and Kondor (2009)).

In terms of actual economic significance, the estimated relationships are also quite sizable. For instance, consider the estimated contemporaneous impact of relative computer taking activity in the same direction (*OFCT*) on triangular arbitrage. The sum of the coefficients is  $-0.03$ , which implies that a one percentage point increase in average AT across all three currency pairs would on average reduce the average number of seconds with a triangular arbitrage in a given minute by 0.03. This sounds like a tiny effect, but the standard deviation of relative computer taking is around 25 percent,<sup>20</sup> suggesting that a one standard deviation increase in AT across all three currency pairs leads to an average reduction of  $3 \times 25 \times 0.03 = 2.25$  seconds of arbitrage opportunities in each minute. On average, in the 3-11am sample, there are only about 2.4 seconds of arbitrage opportunities within each minute, so a reduction of 2.25 seconds seems quite sizeable. Similarly, the causal impact of triangular arbitrage opportunities on AT is fairly large. The standard deviation of triangular arbitrage opportunities is around 3.4,<sup>21</sup> suggesting that a one standard deviation shift in triangular arbitrage leads to a  $3.4 \times 1.3 = 4.42$  percentage point increase in average relative computer taking in each of the three currency pairs.

In summary, both the Granger causality tests and the heteroskedasticity identification approaches point to the same conclusion: AT helps reduce triangular arbitrage opportunities, especially computer taking activity and during intervals of highly correlated computer trading activity, and the presence of triangular arbitrage opportunities leads to an increase in AT.

## 5 Return Serial Correlation and AT

We find that algorithmic traders reduce the number of arbitrage opportunities. However, this is just one example of how computers improve the price discovery process. It is possible that algorithmic traders reduce the number of arbitrage opportunities, but if during times when there are no triangular arbitrage opportunities algorithmic traders behave en masse like the positive-feedback traders of DeLong, Shleifer, Summers, and Waldman (1990) or the chartists described in Froot, Scharfstein, and Stein (1992), or the short-term investors in Vives (1995), then they might cause deviations of prices from fundamental values and thus induce excess volatility. We thus investigate the effect algorithmic traders have on a more general measure of prices not being informationally efficiency: the magnitude of serial autocorrelation in currency returns.<sup>22</sup> In particular,

---

<sup>20</sup>This is roughly the standard deviation of the residuals in the VAR for the relative computer taking equations. The standard deviations in the raw data are a few percentage points higher.

<sup>21</sup>This is again the standard deviation of the residuals in the VAR for triangular arbitrage equation. The standard deviations in the raw triangular arbitrage data is about double this.

<sup>22</sup>Samuelson (1965), Fama (1965) and Fama (1970), among others, show that if prices reflect all public information, then they must follow a martingale process. As a consequence, an informationally efficient price exhibits no serial autocorrelation either positive (momentum) or negative (mean-reversion).

we estimate the serial autocorrelation of 5-second returns over 5-minute intervals.<sup>23</sup> Similar to the evolution of arbitrage opportunities in the market, we show in Figure 5 that the introduction and growth of algorithmic trading coincides with a reduction in the absolute value of serial autocorrelation in the least liquid currency pair, EUR/JPY.

In this section, we adopt the same identification approach as we previously did for triangular arbitrage opportunities, using a 5-minute frequency VAR that allows for both Granger causality tests as well as contemporaneous identification through the heteroskedasticity in the data across the sample period.<sup>24</sup>

In the (structural) VAR analysis in this section, we follow a similar approach to the one used above for triangular arbitrage opportunities. However, since serial correlation is measured separately for each currency pair (unlike triangular arbitrage opportunities), we fit a separate VAR for each currency pair. Algorithmic trading is measured in the same way as before: The fraction of total volume of trade that involves a computer on at least one side of the trade, the relative taking activity of computers, the relative taking activity in the same direction of computers, the relative making activity of computers, and the degree of correlation in computer trading activity.

Let  $|SAC_t^j|$  and  $AT_t^j$  be the five-minute measures of the absolute value of serial autocorrelation in 5-second returns and AT activity, respectively, in currency pair  $j$ ,  $j = 1, 2, 3$ ; as before,  $AT_t^j$  is used to represent either of the five measures of AT activity. Define  $Y_t^j = (|SAC_t^j|, AT_t^j)$ , and the structural form of the system is given by

$$A^j Y_t^j = \Phi^j(L) Y_t^j + \Lambda^j X_{t-1:t-20}^j + \Psi^j G_t + \epsilon_t^j. \quad (3)$$

$A^j$  is the  $2 \times 2$  matrix specifying the contemporaneous effects, normalized such that all diagonal elements are equal to 1.  $X_{t-1:t-20}^j$  includes, for currency pair  $j$ , the sum of the volume of trade over the past 20 minutes and the volatility over the past 20 minutes, calculated as the sum of absolute returns over these 20 minutes.  $G_t$  represent a set of deterministic functions of time  $t$ , capturing individual trends and intra-daily patterns in the variables in  $Y_t$ . In particular,  $G_t = (I_{\{t \in 1st \text{ month of sample}\}}, \dots, I_{\{t \in last \text{ month of sample}\}}, I_{\{t \in 1st \text{ half-hour of day}\}}, \dots, I_{\{t \in last \text{ half-hour of day}\}})$  controls for secular trends and intra-daily patterns in the data by year-month dummy variables and half-hour dummy variables, respectively.<sup>25</sup> The structural shocks  $\epsilon_t^j$  are assumed to be independent of each other and serially uncorrelated at all leads and lags. The number of lags included in the VAR is set to 20.

Table ?? shows the results from both the reduced form and structural identification of equation (3). The

---

<sup>23</sup>Our choice of a 5-second frequency is driven by a trade-off between sampling at a high enough frequency that we estimate the effect algorithmic traders have on prices, but low enough that we have enough transactions to avoid a zero serial autocorrelation bias due to the number of zero-returns. Our results are qualitatively similar when we sample prices at 1-second for the most liquid currency pairs, euro-dollar and dollar-yen, but for the euro-yen the 1-second sampling frequency biases the serial autocorrelation towards zero due to a lack of trading activity at the beginning of the sample.

<sup>24</sup>We estimate the serial correlation of 5-second returns each 5-minute interval.

<sup>25</sup>Replacing the year-month dummy variables with a linear and quadratic trend in  $t$ , yielded very similar estimation results.

results are laid out in the same manner as in Table ??, with the only difference that each currency pair now corresponds to a separate VAR regression. In particular, the first three rows in each panel show the results from a long-run Granger causality test, based on the reduced form of the VAR in equation (3). That is, we test whether the *sum* of the coefficients on the lags of the causing variable is equal to zero. The last two rows in each panel show the results from the contemporaneous identification of the structural form. Identification is again provided by the heteroskedasticity approach of Rigobon (2003) and Rigobon and Sack (2003,2004). The details and validity of this approach for this particular case are provided in the Appendix. The left hand panels show the results for tests of whether algorithmic trading has a causal impact on market depth and the right hand panels show tests of whether market depth has a causal impact on AT activity.

The left panel of Table ?? indicates that an increase in the share of AT participation and relative computer making tends to Granger cause greater price efficiency or lower absolute value of serial autocorrelation in high-frequency exchange rate returns. The impact of relative computer taking appear mixed and in particular a high degree of correlation in computer trading strategies decreases price efficiency in the EUR/USD and USD/JPY currency pairs, although the effect is not statistically significant.

In the right panel, we observe that higher return autocorrelation does not have a strong effect on computer trading activity, in the sense that the contemporaneous coefficients do not show a very strong pattern of statistical significance.

## 6 Conclusion

Using highly-detailed high-frequency trading data for three major exchange rates from 2003 to 2007, we analyze the impact of the growth of algorithmic trading on the spot interdealer foreign exchange market. Algorithmic trading confers a natural speed advantage over human trading, but it also limits the scope of possible trading strategies since any algorithmic strategy must be completely rule-based and pre-programmed. Our results highlight both of these features of algorithmic trading. We show that the rise of algorithmic trading in the foreign exchange market has coincided with a decrease in triangular arbitrage opportunities and a decrease in autocorrelation of high-frequency returns, which is consistent with computers having an enhanced ability to monitor and respond almost instantly to changes in the market. However, our analysis also suggests that the constraint of designing fully systematic (i.e., algorithmic) trading systems leads to less diverse strategies than otherwise, as algorithmic trades (and in the extension, strategies) are found to be more correlated than human ones and this in turn causes higher excess volatility in the EUR/USD and JPY/USD currency pairs, although the effect is not statistically significant.

# Appendix

## A1 Definition of Order Flow and Volume

The transactions data are broken down into categories specifying the “maker” and “taker” of the trades (human or computer), and the direction of the trades (buy or sell the base currency), for a total of eight different combinations. That is, the first transaction category may specify, say, the minute-by-minute volume of trade that results from a computer taker buying the base currency by “hitting” a quote posted by a human maker. We would record this activity as the human-computer buy volume, with the aggressor (taker) of the trade buying the base currency. The human-computer sell volume is defined analogously, as are the other six buy and sell volumes that arise from the remaining combinations of computers and humans acting as makers and takers.

From these eight types of buy and sell volumes, we can construct, for each minute, trading volume and order flow measures for each of the four possible pairs of human and computer makers and takers: human-maker/human-taker ( $HH$ ), computer-maker/human-taker ( $CH$ ), human-maker/computer-taker ( $HC$ ), and computer-maker/computer-taker ( $CC$ ). The *sum* of the buy and sell volumes for each pair gives the volume of trade attributable to that particular combination of maker and taker (denoted as  $Vol(HH)$  or  $Vol(HC)$ , for example). The *difference* between the buy and sell volume for each pair gives the order flow attributable to that maker-taker combination (denoted as  $OF(HH)$  or  $OF(HC)$ , for example). The sum of the four volumes,  $Vol(HH + CH + HC + CC)$ , gives the total volume of trade in the market. The sum of the four order flows,  $OF(HH) + OF(CH) + OF(HC) + OF(CC)$ , gives the total (market-wide) order flow.<sup>26</sup>

Throughout the paper, we use the expression “volume” and “order flow” to refer both to the market-wide volume and order flow and to the volume and order flows from other possible decompositions, with the distinction clearly indicated. Importantly, the data allow us to consider volume and order flow broken down by the type of trader who initiated the trade, human-taker ( $HH + CH$ ) and computer-taker ( $HC + CC$ ); by the type of trader who provided liquidity, human-maker ( $HH + HC$ ) and computer-maker ( $CH + CC$ ); and by whether there was any computer participation ( $HC + CH + CC$ ).

---

<sup>26</sup>There is a very high correlation in this market between trading volume per unit of time and the number of transactions per unit of time, and the ratio between the two does not vary much over our sample. Order flow measures based on amounts transacted and those based on number of trades are therefore very similar.

## A A2 Heteroskedasticity identification

### A.1 Methodology

This section describes in detail the identification and estimation of the simultaneous effect of algorithmic trading on triangular arbitrage opportunities, as captured by equation (1). Restate the structural equation,

$$AY_t = \Phi(L)Y_t + \Lambda X_{t-1:t-20} + \Psi G_t + \epsilon_t,$$

and the reduced form,

$$Y_t = A^{-1}\Phi(L)Y_t + A^{-1}\Lambda X_{t-1:t-15} + A^{-1}\Psi G_t + A^{-1}\epsilon_t.$$

Let  $\Omega_s$  be the variance-covariance matrix of the reduced form errors in variance regime  $s$ ,  $s = 1, 2$ , which can be directly estimated by  $\hat{\Omega}_s = \frac{1}{T_s} \sum_{t \in s} u_t u_t'$ , where  $u_t$  are the reduced form residuals. Let  $\Omega_{\epsilon,s}$  be the diagonal variance-covariance matrix of the structural errors and the following moment conditions hold,

$$A\Omega_s A' = \Omega_{\epsilon,s}. \tag{4}$$

The parameters in  $A$  and  $\Omega_{\epsilon,s}$  can then be estimated with GMM, using estimates of  $\Omega_s$  from the reduced form equation. Identification is achieved as long as the covariance matrices constitute a system of equations that is linearly independent. In the case of two regimes, the system is exactly identified. Standard errors are calculated via bootstrapping, resampling daily blocks of data to control for any remaining intra-daily pattern, and using 200 repetitions.

### A.2 Covariance matrix estimates in the triangular arbitrage case

The heteroskedasticity identification approach requires there to exist two linearly independent variance regimes. Table C1 shows the estimates of the covariance matrix for the reduced form VAR residuals for  $Y_t = (Arb_t, AT_t^1, AT_t^2, AT_t^3)$ , across the first and second half of the sample. As is clear, there is a strong increase in the variance of algorithmic trading, for all three measures used. In contrast, the variance in triangular arbitrage opportunities is almost constant across the two subsamples. It is thus immediately clear that the change in covariance matrix between the two sub samples is not proportional.

### A.3 Covariance matrix estimates in the liquidity case

Table ?? shows the estimates of the covariance matrices, for each currency pair separately, for the VAR residuals with AT and liquidity. Again, it is clear that there is not a proportional change in the covariance matrix between the first and second half of the sample.

## A3 How Correlated Are Algorithmic Trades and Strategies?

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

$$Prob(HH) = (1 - \alpha_m)(1 - \alpha_t)$$

$$Prob(HC) = (1 - \alpha_m)\alpha_t$$

$$Prob(CH) = \alpha_m(1 - \alpha_t)$$

$$Prob(CC) = \alpha_m\alpha_t.$$

These probabilities yield the following identity,

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

which can be re-written as,

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

We label the first ratio,  $RH \equiv \frac{Prob(HH)}{Prob(CH)}$ , the “human-taker” ratio and the second ratio,  $RC \equiv \frac{Prob(HC)}{Prob(CC)}$ , the “computer-taker” ratio. In a world with more human traders (both makers and takers) than computer traders, each of these ratios will be greater than one, because  $Prob(HH) > Prob(CH)$  and  $Prob(HC) > Prob(CC)$ ; i.e., computers take liquidity more from humans than from other computers, and humans take liquidity more from humans than from computers. However, under the baseline assumptions of our random-

matching model, the identity shown above states that the ratio of ratios,  $R \equiv \frac{RC}{RH}$ , will be equal to one. In other words, humans will take liquidity from other humans in a similar proportion that computers take liquidity from humans.

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

To explicitly take into account the sign of trades, we amend the benchmark model as follows: we assume that the probability of the taker buying an asset is  $\alpha_B$  and the probability of the taker selling is  $1 - \alpha_B$ . We can then write the probability of the following eight events (assuming each event is independent):

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

These probabilities yield the following identities,

$$Prob(HH^B) \times Prob(CC^B) \equiv Prob(HC^B) \times Prob(CH^B)$$

$$(1 - \alpha_m)(1 - \alpha_t)\alpha_m\alpha_t\alpha_B\alpha_B \equiv (1 - \alpha_m)\alpha_t\alpha_B\alpha_m(1 - \alpha_t)\alpha_B$$

and

$$Prob(HH^S) \times Prob(CC^S) \equiv Prob(HC^S) \times Prob(CH^S)$$

$$(1 - \alpha_m)(1 - \alpha_t)(1 - \alpha_B)\alpha_m\alpha_t\alpha_B(1 - \alpha_B) \equiv (1 - \alpha_m)\alpha_t\alpha_B(1 - \alpha_B)\alpha_m(1 - \alpha_t)\alpha_B(1 - \alpha_B)$$



which can be re-written as,

$$\begin{aligned} \frac{Prob(HH^B)}{Prob(CH^B)} &\equiv \frac{Prob(HC^B)}{Prob(CC^B)} \\ \frac{(1-\alpha_m)(1-\alpha_t)\alpha_B}{\alpha_m(1-\alpha_t)\alpha_B} &\equiv \frac{(1-\alpha_m)\alpha_t\alpha_B}{\alpha_m\alpha_t\alpha_B} \\ &\text{and} \\ \frac{Prob(HH^S)}{Prob(CH^S)} &\equiv \frac{Prob(HC^S)}{Prob(CC^S)} \\ \frac{(1-\alpha_m)(1-\alpha_t)(1-\alpha_B)}{\alpha_m(1-\alpha_t)(1-\alpha_B)} &\equiv \frac{(1-\alpha_m)\alpha_t(1-\alpha_B)}{\alpha_m\alpha_t(1-\alpha_B)} \end{aligned}$$

We label the ratios,  $RH^B \equiv \frac{Prob(HH^B)}{Prob(CH^B)}$ , the “human-taker-buyer” ratio,  $RC^B \equiv \frac{Prob(HC^B)}{Prob(CC^B)}$ , the “computer-taker-buyer” ratio,  $RH^S \equiv \frac{Prob(HH^S)}{Prob(CH^S)}$ , the “human-taker-seller” ratio, and  $RC^S \equiv \frac{Prob(HC^S)}{Prob(CC^S)}$ , the “computer-taker-seller” ratio.

In a world with more human traders (both makers and takers) than computer traders, each of these ratios will be greater than one, because  $Prob(HH^B) > Prob(CH^B)$ ,  $Prob(HH^S) > Prob(CH^S)$ ,  $Prob(HC^B) > Prob(CC^B)$ , and  $Prob(HC^S) > Prob(CC^S)$ . That is, computers take liquidity more from humans than from other computers, and humans take liquidity more from humans than from computers. However, under the baseline assumptions of our random-matching model, the identity shown above states that the ratio of ratios,  $R^B \equiv \frac{RC^B}{RH^B}$ , will be equal to one, and  $R^S \equiv \frac{RC^S}{RH^S}$ , will also be equal to one.

Under the assumption that potential human-taker-buyers are randomly matched with potential human-maker-sellers and human-taker-sellers are randomly matched with potential human-maker-buyers, i.e., that the probability of a human-maker-seller/human-taker-buyer trade is equal to the one predicted by our model,  $Prob(HH^B) = (1-\alpha_m)(1-\alpha_t)\alpha_B$ , and the probability of a human-maker-buyer/human-taker-seller trade is equal to the one predicted by our model,  $Prob(HH^S) = (1-\alpha_m)(1-\alpha_t)(1-\alpha_B)$ , we can now derive implications from observations of  $R^B$  and  $R^S$  our ratio of ratios. In particular, finding  $R^B > 1$  must imply that algorithmic strategies of buyers are more correlated than what our random matching model implies. In other words, for  $R^B > 1$  we must observe that either computers trade with each other less than expected when they are buying ( $Prob(CC^B) < \alpha_m\alpha_t\alpha_B$ ) or that computers trade with humans more than expected when they are buying (either  $Prob(CH^B) > \alpha_m(1-\alpha_t)\alpha_B$  or  $Prob(HC^B) > (1-\alpha_m)\alpha_t\alpha_B$ ). Symmetrically, for  $R^S > 1$  we must observe that either computers trade with each other less than expected when they are selling ( $Prob(CC^S) < \alpha_m\alpha_t(1-\alpha_B)$ ) or that computers trade with humans more than expected when they are selling (either  $Prob(CH^S) > \alpha_m(1-\alpha_t)(1-\alpha_B)$  or  $Prob(HC^S) > (1-\alpha_m)\alpha_t(1-\alpha_B)$ ).

Table A1: Correlation among algorithmic trading strategies.

The table reports estimates of the relative degree to which computers trade with each other compared to how much they trade with humans, based on the benchmark model described in the main text. In particular, we report mean estimates of the log of the sell and buy ratios  $R^S = RC^S / RH^S$  and  $R^B = RC^B / RH^B$ , measured at the 1-minute, 5-minute, and daily frequency, with standard errors shown in parentheses below the estimates.  $\ln(R)$ ,  $\ln(R^S)$ ,  $\ln(R^B) > 0$  ( $R$ ,  $R^S$ ,  $R^B > 1$ ) indicates that computers trade less with each other than random matching would predict. The percent of observations where  $\ln(R)$ ,  $\ln(R^S)$ ,  $\ln(R^B) > 0$  is also reported along with the number of non-missing observations and the total number of observations, at each frequency. The \*\*\*, \*\*, and \* represent a statistically significant deviation from zero at the 1, 5, and 10 percent level, respectively. The sample period is January 2006 to December 2007.

	1-min data			5-min data			Daily data		
	$\ln(R)$	$\ln(R^S)$	$\ln(R^B)$	$\ln(R)$	$\ln(R^S)$	$\ln(R^B)$	$\ln(R)$	$\ln(R^S)$	$\ln(R^B)$
Mean	0.2955***	0.1154***	0.0983***	0.4466***	0.2937***	0.2913***	0.3948***	0.3917***	0.3914***
(std. err.)	(0.0032)	(0.0043)	(0.0043)	(0.0036)	(0.0046)	(0.0047)	(0.0051)	(0.006)	(0.006)
Percent of obs.>0	0.624	0.546	0.54	0.759	0.658	0.657	1	1	0.996
No. of non-missing obs.	130333	83404	81556	42966	38381	37947	498	498	498
Total no. of obs.	239040	239040	239040	47808	47808	47808	498	498	498
Mean	0.3809***	0.201***	0.1992***	0.4832***	0.2906***	0.2951***	0.3828***	0.381***	0.375***
(std. err.)	(0.0038)	(0.0053)	(0.0054)	(0.0043)	(0.0055)	(0.0056)	(0.0059)	(0.0071)	(0.007)
Percent of obs.>0	0.634	0.561	0.561	0.74	0.638	0.64	0.998	0.992	0.988
No. of non-missing obs.	103885	57098	57681	40800	33950	34183	498	498	498
Total no. of obs.	239040	239040	239040	47808	47808	47808	498	498	498
Mean	0.7748***	0.5794***	0.5652***	0.8172***	0.6902***	0.6883***	0.6521***	0.6469***	0.6499***
(std. err.)	(0.0052)	(0.0078)	(0.008)	(0.0054)	(0.0071)	(0.0071)	(0.0096)	(0.0103)	(0.0114)
Percent of obs.>0	0.72	0.658	0.651	0.804	0.736	0.734	1	0.994	0.982
No. of non-missing obs.	62323	27785	26750	36055	26287	25866	498	498	498
Total no. of obs.	239040	239040	239040	47808	47808	47808	498	498	498

Table B1: Correlation among algorithmic trading strategies.

The table reports estimates of the relative degree to which computers trade with each other compared to how much they trade with humans, based on the benchmark model described in the main text. In particular, we report mean estimates of the Taylor expansion approximation of the log of the ratios  $R = RC/RH$ ,  $R^S = RC^S/RH^S$  and  $R^B = RC^B/RH^B$ , around the mean values for volume measured at the 1-minute, 5-minute, and daily frequency, with standard errors shown in parentheses below the estimates.  $\log(R^S), \log(R^B) > 0$  ( $R^S, R^B > 1$ ) indicates that computers trade less with each other than random matching would predict. The percent of observations where  $\log(R^S), \log(R^B) > 0$  is also reported along with the number of non-missing observations and the total number of observations, at each frequency. The \*\*\*, \*\*, and \* represent a statistically significant deviation from zero at the 1, 5, and 10 percent level, respectively. The sample period is January 2006 to December 2007.

	1-min data			5-min data			Daily data		
	$\ln(R)$	$\ln(R^S)$	$\ln(R^B)$	$\ln(R)$	$\ln(R^S)$	$\ln(R^B)$	$\ln(R)$	$\ln(R^S)$	$\ln(R^B)$
Mean	0.4635***	0.4624***	0.463***	0.4694***	0.3734***	0.3753***	0.3950***	0.3915***	0.3915***
(std. err.)	(0.002)	(0.0025)	(0.0025)	(0.0032)	(0.0038)	(0.0038)	(0.0049)	(0.0059)	(0.0058)
Percent of obs.>0	0.737	0.711	0.709	0.781	0.713	0.714	1	1	1
No. of non-missing obs.	239040	239040	239040	47808	47808	47808	498	498	498
Total no. of obs.	239040	239040	239040	47808	47808	47808	498	498	498
Mean	0.5133***	0.4881***	0.4887***	0.5165***	0.4001***	0.4026***	0.3822***	0.3802***	0.3751***
(std. err.)	(0.0024)	(0.0031)	(0.0031)	(0.0037)	(0.0043)	(0.0044)	(0.0057)	(0.0068)	(0.0069)
Percent of obs.>0	0.726	0.697	0.695	0.769	0.703	0.704	1	1	0.988
No. of non-missing obs.	239040	239040	239040	47808	47808	47808	498	498	498
Total no. of obs.	239040	239040	239040	47808	47808	47808	498	498	498
Mean	0.7329***	0.7014***	0.7086***	0.7967***	0.7113***	0.7175***	0.6523***	0.6461***	0.6493***
(std. err.)	(0.0032)	(0.0048)	(0.0048)	(0.0044)	(0.0051)	(0.0052)	(0.0095)	(0.01)	(0.0112)
Percent of obs.>0	0.755	0.732	0.726	0.817	0.768	0.766	1	1	0.982
No. of non-missing obs.	239040	239040	239040	47808	47808	47808	498	498	498
Total no. of obs.	239040	239040	239040	47808	47808	47808	498	498	498

Table C1: Covariances from residuals of the bivariate VAR with triangular arbitrage and algorithmic trading activity.

The table shows the covariance matrices for the residuals from the reduced form VAR in equation (2), which is estimated separately for each of the following measures of algorithmic trading activity averaged across currency pairs: Overall AT participation ( $VAT$ ), AT taking participation ( $VCt$ ), AT making participation ( $VCm$ ), Relative Computer Taking ( $OFCT$ ), and the natural logarithm of AT trade correlation ( $\ln(R)$ ). The VAR is estimated using 1-minute data for the full sample from 2003 to 2007, spanning 1067 days. The residuals are split up into two sub samples, covering the first 534 days and last 533 days, respectively, and the covariance matrices are calculated separately for these two sub samples. The final column in the table shows the ratios between the estimates from the second and the first sub samples.

	First half of sample	Second half of sample	Ratio (second/first)
$VAT$			
$Var(Arb)$	11.52	11.07	0.96
$Var(VAT)$	119.26	237.65	1.99
$Cov(Arb, VAT)$	5.32	3.31	0.62
$VCt$			
$Var(Arb)$	11.52	11.07	0.96
$Var(VCt)$	76.14	223.00	2.93
$Cov(Arb, VCt)$	5.06	3.86	0.76
$VCm$			
$Var(Arb)$	11.52	11.07	0.96
$Var(VCm)$	51.79	151.30	2.92
$Cov(Arb, VCm)$	0.42	-0.01	-0.02
$OFCT$			
$Var(Arb)$	11.50	11.05	0.96
$Var(OFCT)$	196.80	360.28	1.83
$Cov(Arb, OFCT)$	9.68	4.18	0.43
$\ln(R)$			
$Var(Arb)$	17.17	10.32	0.60
$Var(\log(R))$	1.32	1.02	0.78
$Cov(Arb, \log(R))$	0.64	0.15	0.24

Table C2: Covariances from residuals of the bivariate VAR with high-frequency autocorrelation and algorithmic trading activity.

The table shows the covariance matrices for the residuals from the reduced form VAR in equation (??), which is estimated separately for each currency pair and each of the following measures of algorithmic trading activity: Overall AT participation ( $VAT$ ), AT taking participation ( $VCt$ ), AT making participation ( $VCm$ ), Relative Computer Taking ( $OFcT$ ), and the natural logarithm of AT trade correlation ( $\ln(R)$ ). The VAR is estimated using 5-minute data for the full sample from 2003 to 2007, spanning 1067 days. The residuals are split up into two sub samples, covering the first 534 days and last 533 days, respectively, and the covariance matrices are calculated separately for these two sub samples. The final column in the table shows the ratios between the estimates from the second and the first sub samples.

	First half of sample			Second half of sample			Ratio (second/first)		
	USD/EUR	JPY/USD	JPY/EUR	USD/EUR	JPY/USD	JPY/EUR	USD/EUR	JPY/USD	JPY/EUR
$Var(ac)$	111.98	136.27	180.68	127.15	118.82	125.22	1.14	0.87	0.69
$Var(VAT)$	12.34	50.39	210.03	91.66	160.78	258.44	7.43	3.19	1.23
$Cov(ac, VAT)$	-0.22	-1.06	2.01	-3.95	-3.70	-2.31	18.02	3.49	-1.15
	$VCt$								
$Var(ac)$	111.97	136.30	180.70	127.20	118.85	125.22	1.14	0.87	0.69
$Var(VCt)$	4.84	23.11	143.71	66.35	131.13	277.61	13.72	5.67	1.93
$Cov(ac, VCt)$	0.03	-0.60	-1.70	-1.57	-1.16	-1.63	-57.73	1.94	0.96
	$VCm$								
$Var(ac)$	111.97	136.28	180.67	127.16	118.83	125.21	1.14	0.87	0.69
$Var(VCm)$	6.32	25.92	96.48	52.81	101.01	203.51	8.35	3.90	2.11
$Cov(ac, VCm)$	-0.27	-0.65	3.32	-3.54	-3.63	-1.00	12.98	5.60	-0.30
	$OFcT$								
$Var(ac)$	111.89	136.10	180.31	127.22	118.90	125.09	1.14	0.87	0.69
$Var(OFcT)$	252.23	386.55	587.54	620.48	664.80	791.08	2.46	1.72	1.35
$Cov(ac, OFcT)$	0.02	2.64	2.37	1.47	0.24	1.17	67.63	0.09	0.49
	$\ln(R)$								
$Var(ac)$	98.85	108.74	137.68	124.19	113.63	116.48	1.26	1.04	0.85
$Var(\log(R))$	0.95	1.10	1.19	0.58	0.77	1.02	0.61	0.70	0.85
$Cov(ac, \log(R))$	-0.04	0.01	-0.15	-0.25	-0.21	-0.16	6.69	-21.41	1.07

Table C3: Covariances from residuals of the trivariate VAR with triangular arbitrage, AT activity and AT trade correlation.

The table shows the covariance matrices for the residuals from the reduced form VAR in equation (??), which is estimated using the measure of triangular arbitrage ( $arb$ ), Relative Computer Taking ( $OFCT$ ), and the natural logarithm of AT trade correlation ( $\ln(R)$ ), with the latter two variables averaged across currency pairs. The VAR is estimated using 1-minute data for the full sample from 2003 to 2007, spanning 1067 days. The residuals are split up into two sub samples, covering the first 534 days and last 533 days, respectively, and the covariance matrices are calculated separately for these two sub samples. The final column in the table shows the ratios between the estimates from the second and the first sub samples.

	First half of sample	Second half of sample	Ratio (second/first)
$Var(Arb)$	17.17	10.31	0.60
$Var(OFCT)$	270.10	296.43	1.10
$Var(\ln(R))$	1.32	1.02	0.78
$Cov(Arb, OFCT)$	9.31	2.72	0.29
$Cov(Arb, \ln(R))$	0.64	0.15	0.24
$Cov(OFCT, \ln(R))$	2.86	1.21	0.42

Table C4: Covariances from residuals of the trivariate VAR with high-frequency autocorrelation, AT activity and AT trade correlation.

The table shows the covariance matrices for the residuals from the reduced form VAR in equation (??), which is estimated separately for each currency pair using the measure of autocorrelation ( $ac$ ), Relative Computer Taking ( $OFCT$ ), and the natural logarithm of AT trade correlation ( $\ln(R)$ ). The VAR is estimated using 5-minute data for the full sample from 2003 to 2007, spanning 1067 days. The residuals are split up into two sub samples, covering the first 534 days and last 533 days, respectively, and the covariance matrices are calculated separately for these two sub samples. The final column in the table shows the ratios between the estimates from the second and the first sub samples.

	First half of sample			Second half of sample			Ratio (second/first)		
	USD/EUR	JPY/USD	JPY/EUR	USD/EUR	JPY/USD	JPY/EUR	USD/EUR	JPY/USD	JPY/EUR
$Var(ac)$	98.87	108.81	137.69	124.08	113.60	116.46	1.26	1.04	0.85
$Var(VCm)$	16.76	40.43	96.32	51.74	91.60	141.82	3.09	2.27	1.47
$Var(\log(R))$	0.95	1.10	1.19	0.58	0.77	1.02	0.61	0.70	0.85
$Cov(ac, VCm)$	-0.48	0.75	2.40	-3.27	-2.98	0.39	6.89	-3.98	0.16
$Cov(ac, \log(R))$	-0.04	0.01	-0.15	-0.24	-0.21	-0.16	6.20	-35.20	1.08
$Cov(VCm, \log(R))$	0.94	1.48	1.82	0.53	0.80	0.33	0.56	0.54	0.18

## References

- [1] 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.
- [2] 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.
- [3] Andersen, T.G., T. Bollerslev, F.X. Diebold, and C. Vega, 2007. Real-time price discovery in global stock, bond, and foreign exchange markets, *Journal of International Economics* 73, 251-277.
- [4] Bandi, F.M., and J.R. Russell, 2006, Separating Microstructure Noise from Volatility, *Journal of Financial Economics*, 79, 655-692.
- [5] Bertsimas, D., and Lo, A., 1998, Optimal Control of Execution Costs, *Journal of Financial Markets* 1, 1-50.
- [6] Biais, B., Foucault, T., and Moinas S., 2011, Equilibrium Algorithmic Trading, Working Paper, Toulouse School of Economics.
- [7] Biais, B., Hillion, P., and Spatt, C., 1995, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *Journal of Finance*, 50, 1655-1689.
- [8] Biais, B., and P. Woolley, High Frequency Trading,
- [9] Bloomfield, R., O'Hara, M., and Saar, G., 2009, How noise trading affects markets: an experimental analysis, *Review of Financial Studies*, 22, 2275-2302.
- [10] Brogaard, J. A., 2010, High Frequency Trading and Its Impact on Market Quality, working paper, Northwestern University.
- [11] Chaboud, A., Chiquoine, B., Hjalmarsson, E., Loretan, M., 2010, Frequency of Observations and the Estimation of Integrated Volatility in Deep and Liquid Financial Markets, *Journal of Empirical Finance*, 17, 212-240.
- [12] Chaboud, A., Chernenko, S., and Wright, J., 2008, Trading Activity and Macroeconomic Announcements in High-Frequency Exchange Rate Data, *Journal of the European Economic Association*, 6, 589-596.
- [13] DeLong, B., Shleifer, A., Summers, L., Waldmann, R.J., 1990, Positive Feedback Investment Strategies and Destabilizing Rational Speculation, *Journal of Finance*, 45, 379-395.



- [14] Easley, D., and O'Hara, M., 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics*, 19, 69-90.
- [15] Evans, M., and R. Lyons, 2008. How is Macro News Transmitted to Exchange Rates?, *Journal of Financial Economics* 88, 26-50.
- [16] Fama, E., 1965, The Behavior of Stock Market Prices, *Journal of Business*, 38.
- [17] Fama, E., 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance*, 25.
- [18] Foucault, T., 2012, Algorithmic Trading: Issues and Preliminary Evidence, Market Microstructure Confronting Many Viewpoints, John Wiley & Sons.
- [19] Foucault, T., Kadan, O., and Kandel, E., 2009, Liquidity Cycles and Make/Take Fees in Electronic Markets, working paper, Washington University in St. Louis.
- [20] Foucault, T., and Menkveld, A., 2008, Competition for Order Flow and Smart Order Routing Systems, *Journal of Finance*, 63, 119-158.
- [21] Froot, K., Scharfstein, D. S., Stein, J.C., 1992, Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation, *Journal of Finance*, 47, 1461-1484.
- [22] Hasbrouck, J., 1991a. Measuring the information content of stock trades, *Journal of Finance* 46, 179-207.
- [23] Hasbrouck, J., 1991b. The summary informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 4, 571-595.
- [24] Hasbrouck, J., 1996. Order characteristics and stock price evolution: An application to program trading, *Journal of Financial Economics* 41, 129-149.
- [25] He, H., and J. Wang, 1995, Differential information and dynamic behavior of stock trading volume, *Review of Financial Studies*, 8, 919-972.
- [26] Hendershott, T., C.M. Jones, and A.J. Menkveld, 2011. Does algorithmic trading improve liquidity?, *Journal of Finance*, 66, 1-33.
- [27] Hendershott, T., and Riordan, R., 2009, Algorithmic Trading and Information, working paper, University of California at Berkeley.
- [28] Hoffman, P., 2012, A Dynamic Limit Order Market with Fast and Slow Traders, Working Paper.

- [29] Jarrow, R., and P. Protter, 2011, A Dysfunctional Role of High Frequency Trading in Electronic Markets, Cornell University Working Paper.
- [30] Khandani, A., and Lo, A., 2007, What Happened to the Quants in August 2007?, *Journal of Investment Management* 5, 5-54.
- [31] Khandani, A., and Lo, A., 2011, What Happened to the Quants in August 2007? Evidence from Factors and Transactions Data, *Journal of Financial Markets* 14, 1-46.
- [32] Kondor, p., 2009, Risk in Dynamic Arbitrage: Price Effects of Convergence Trading, *Journal of Finance*, 64, 631-655.
- [33] Kozhan, R., and W. Wah Tham, 2012, Execution Risk in High-Frequency Arbitrage, *Management Science*, forthcoming.
- [34] Lo, A., Repin, D.V., and Steenbarger, B.N., 2005., Fear and Greed in Financial Markets: A Clinical Study of Day-Traders, *American Economic Review* 95, 352–359.
- [35] Martinez, V., and I. Roşu, 2011, High Frequency Traders, News and Volatility, Working Paper.
- [36] Oehmke, M., 2009, Gradual Arbitrage, Working Paper, Columbia University.
- [37] Pasquariello, P., and Vega, C., 2007, Informed and Strategic Order Flow in the Bond Markets, *Review of Financial Studies*, 20, 1975-2019.
- [38] Samuelson, P. A., 1965, Proof that Properly Anticipated Prices Fluctuate Randomly, *Industrial Management Review*.
- [39] Stein, J., Presidential Address: Sophisticated Investors and Market Efficiency, *Journal of Finance*, 64, 1571-1548.
- [40] Stock, J.H., J.H. Wright, and M. Yogo, 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics*, 20, 518-529.
- [41] 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.
- [42] Zhang, X.F., 2010, The Effect of High-Frequency Trading on Stock Volatility and Price Discovery, working paper, Yale University.

Table 3: Correlation among algorithmic trading strategies.

The table reports estimates of the relative degree to which computers trade with each other compared to how much they trade with humans, based on the benchmark model described in the main text. In particular, we report mean estimates of the log of the sell and buy ratios  $R^S = RC^S / RH^S$  and  $R^B = RC^B / RH^B$ , measured at the 1-minute, 5-minute, and daily frequency, with standard errors shown in parentheses below the estimates.  $\ln(R)$ ,  $\ln(R^S)$ ,  $\ln(R^B) > 0$  ( $R, R^S, R^B > 1$ ) indicates that computers trade less with each other than random matching would predict. The percent of observations where  $\ln(R)$ ,  $\ln(R^S)$ ,  $\ln(R^B) > 0$  is also reported along with the number of non-missing observations and the total number of observations, at each frequency. The \*\*\*, \*\*, and \* represent a statistically significant deviation from zero at the 1, 5, and 10 percent level, respectively. The sample period is September 2003 to December 2007.

	1-min data			5-min data			Daily data		
	$\ln(R)$	$\ln(R^S)$	$\ln(R^B)$	$\ln(R)$	$\ln(R^S)$	$\ln(R^B)$	$\ln(R)$	$\ln(R^S)$	$\ln(R^B)$
Mean	0.2216***	0.0545***	0.0399***	0.3672***	0.2116***	0.2074***	0.531***	0.4896***	0.4993***
(std. err.)	(0.0031)	(0.0042)	(0.0042)	(0.0036)	(0.0046)	(0.0047)	(0.0118)	(0.0122)	(0.0118)
Percent of obs.>0	0.599	0.527	0.522	0.722	0.627	0.626	0.99	0.975	0.974
No. of non-missing obs.	143539	89960	87597	52174	44366	43609	881	847	855
Total no. of obs.	512640	512640	512640	102528	102528	102528	1068	1068	1068
					USD/EUR				
Mean	0.3106***	0.1535***	0.1513***	0.3923***	0.2119***	0.2111***	0.5846***	0.5755***	0.5398***
(std. err.)	(0.0037)	(0.0052)	(0.0052)	(0.0043)	(0.0054)	(0.0054)	(0.0131)	(0.0143)	(0.0133)
Percent of obs.>0	0.611	0.545	0.545	0.703	0.609	0.61	0.99	0.969	0.971
No. of non-missing obs.	114538	61048	61683	49980	39077	39413	954	939	923
Total no. of obs.	512640	512640	512640	102528	102528	102528	1068	1068	1068
					JPY/USD				
Mean	0.6984 **	0.531***	0.5196***	0.6873***	0.584***	0.5837***	0.8129***	0.7736***	0.7376***
(std. err.)	(0.0049)	(0.0074)	(0.0076)	(0.0051)	(0.0066)	(0.0067)	(0.0173)	(0.0171)	(0.0163)
Percent of obs.>0	0.696	0.643	0.636	0.758	0.7	0.698	0.984	0.975	0.965
No. of non-missing obs.	71810	30571	29346	45889	31648	31028	988	952	943
Total no. of obs.	512640	512640	512640	102528	102528	102528	1068	1068	1068



Table 4: Triangular arbitrage and algorithmic trading.

We report tests of whether algorithmic trading activity has a causal impact on triangular arbitrage (left hand panels) and whether triangular arbitrage has a causal impact on algorithmic trading activity (right hand panels). Results are presented separately for each of the following measures of algorithmic trading activity averaged across currency pairs: Overall AT participation (*VAT*), AT taking participation (*VCT*), AT making participation (*VCm*), Relative Computer Taking (*OFCT*), and the natural logarithm of AT trade correlation ( $\ln(R)$ ). All results are based on 1-minute data covering the full sample period from 2003 to 2007. The first seven rows in each panel presents the results from three different Granger causality tests, based on the reduced form VAR in equation (2). In particular, the first two rows in each panel reports the coefficient estimate and standard error of the first lag-coefficient for the causing variable. Rows three to five report the sum of the lag-coefficients for the causing variable, along with the corresponding Wald  $\chi^2$ -statistic and p-value for the null hypothesis that this sum is equal to zero, respectively. The sixth and seventh row report the Wald  $\chi^2$ -statistic and p-value for the null hypothesis that the coefficients on all lags of the causing variable are jointly equal to zero. The following row, labeled Contemp. coeff., presents the point estimate of the contemporaneous impact of the causing variable in the structural VAR in equation (1), based on the heteroskedasticity identification scheme described in the main text; the Newey-West standard error is presented below in parentheses. The last five rows in each panel show, respectively, the total number of observations available for estimation in the full sample, the number of observations available in each of the two sub samples used in the heteroskedasticity identification scheme, and number of different days that are included in each of these two sub samples. The \*\*\*, \*\*, and \* represent a statistically significant deviation from zero at the 1, 5, and 10 percent level, respectively.

Tests of <i>VAT</i> Causing Triangular Arbitrage		Tests of Triangular Arbitrage Causing <i>VAT</i>	
Sum of coeffs. on <i>VAT</i> lags	-0.0027***	Sum of coeffs. on <i>arb</i> lags	0.0140***
$\chi^2$ (Sum = 0)	7.7079	$\chi^2$ (Sum = 0)	19.5814
p-value	0.0055	p-value	0.0000
$\chi^2_0$ (All coeffs. on <i>VAT</i> lags = 0)	91.4066***	$\chi^2_0$ (All coeffs. on <i>arb</i> lags = 0)	83.0479***
p-value	0.0000	p-value	0.0000
Contemp. coeff.	-0.0145***	Contemp. coeff.	0.6072***
(std. err.)	(0.0040)	(std. err.)	(0.0591)
No. of obs. in 1st sub sample	512016	No. of obs.	512016
No. of obs. in 2nd sub sample	256246	No. of obs. in 1st sub sample	256246
No. of obs. in 2nd sub sample	255770	No. of obs. in 2nd sub sample	255770
No. of unique days in 1st sub sample	534	No. of unique days in 1st sub sample	534
No. of unique days in 2nd sub sample	533	No. of unique days in 2nd sub sample	533
Tests of <i>VCT</i> Causing Triangular Arbitrage		Tests of Triangular Arbitrage Causing <i>VCT</i>	
Sum of coeffs. on <i>VCT</i> lags	-0.0061***	Sum of coeffs. on <i>arb</i> lags	0.0219***
$\chi^2$ (Sum = 0)	25.2501	$\chi^2$ (Sum = 0)	56.9040
p-value	0.0000	p-value	0.0000
$\chi^2_0$ (All coeffs. on <i>VCT</i> lags = 0)	149.5884***	$\chi^2_0$ (All coeffs. on <i>arb</i> lags = 0)	118.9288***
p-value	0.0000	p-value	0.0000
Contemp. coeff.	-0.0067**	Contemp. coeff.	0.4817***
(std. err.)	(0.0027)	(std. err.)	(0.0331)
No. of obs. in 1st sub sample	512016	No. of obs.	512016
No. of obs. in 2nd sub sample	256246	No. of obs. in 1st sub sample	256246
No. of obs. in 2nd sub sample	255770	No. of obs. in 2nd sub sample	255770
No. of unique days in 1st sub sample	534	No. of unique days in 1st sub sample	534
No. of unique days in 2nd sub sample	533	No. of unique days in 2nd sub sample	533

Table 4: Triangular arbitrage and algorithmic trading. (cont.)

Tests of $VCm$ Causing Triangular Arbitrage		Tests of Triangular Arbitrage Causing $VCm$	
Sum of coeffs. on $VCm$ lags	0.0049***	Sum of coeffs. on $arb$ lags	-0.0022
$\chi^2_1$ (Sum = 0)	14.4351	$\chi^2_1$ (Sum = 0)	0.8233
p-value	0.0001	p-value	0.3642
$\chi^2_0$ (All coeffs. on $VCm$ lags = 0)	46.5108***	$\chi^2_0$ (All coeffs. on $arb$ lags = 0)	31.0039*
p-value	0.0007	p-value	0.0551
Contemp. coeff.	-0.0040***	Contemp. coeff.	0.0545***
(std. err.)	(0.0011)	(std. err.)	(0.0089)
No. of obs.	512016	No. of obs.	512016
No. of obs. in 1st sub sample	256246	No. of obs. in 1st sub sample	256246
No. of obs. in 2nd sub sample	255770	No. of obs. in 2nd sub sample	255770
No. of unique days in 1st sub sample	534	No. of unique days in 1st sub sample	534
No. of unique days in 2nd sub sample	533	No. of unique days in 2nd sub sample	533
Tests of $OFct$ Causing Triangular Arbitrage		Tests of Triangular Arbitrage Causing $OFct$	
Sum of coeffs. on $OFct$ lags	-0.0117***	Sum of coeffs. on $arb$ lags	0.0345***
$\chi^2_1$ (Sum = 0)	142.4661	$\chi^2_1$ (Sum = 0)	75.9510
p-value	0.0000	p-value	0.0000
$\chi^2_0$ (All coeffs. on $OFct$ lags = 0)	317.3994***	$\chi^2_0$ (All coeffs. on $arb$ lags = 0)	459.3056***
p-value	0.0000	p-value	0.0000
Contemp. coeff.	-0.0288***	Contemp. coeff.	1.3037***
(std. err.)	(0.0055)	(std. err.)	(0.1241)
No. of obs.	511688	No. of obs.	511688
No. of obs. in 1st sub sample	255972	No. of obs. in 1st sub sample	255972
No. of obs. in 2nd sub sample	255716	No. of obs. in 2nd sub sample	255716
No. of unique days in 1st sub sample	534	No. of unique days in 1st sub sample	534
No. of unique days in 2nd sub sample	533	No. of unique days in 2nd sub sample	533
Tests of $\ln(R)$ Causing Triangular Arbitrage		Tests of Triangular Arbitrage Causing $\ln(R)$	
Sum of coeffs. on $\ln(R)$ lags	-0.0437	Sum of coeffs. on $arb$ lags	0.0011***
$\chi^2_1$ (Sum = 0)	1.6112	$\chi^2_1$ (Sum = 0)	7.6869
p-value	0.2043	p-value	0.0056
$\chi^2_0$ (All coeffs. on $\ln(R)$ lags = 0)	24.8891	$\chi^2_0$ (All coeffs. on $arb$ lags = 0)	48.0498***
p-value	0.2057	p-value	0.0004
Contemp. coeff.	-0.8782**	Contemp. coeff.	0.1013**
(std. err.)	(0.4413)	(std. err.)	(0.0400)
No. of obs.	147114	No. of obs.	147114
No. of obs. in 1st sub sample	8530	No. of obs. in 1st sub sample	8530
No. of obs. in 2nd sub sample	138584	No. of obs. in 2nd sub sample	138584
No. of unique days in 1st sub sample	184	No. of unique days in 1st sub sample	184
No. of unique days in 2nd sub sample	533	No. of unique days in 2nd sub sample	533

Table 5: High-frequency autocorrelation and algorithmic trading.

We report tests of whether algorithmic trading activity has a causal impact on autocorrelation (left hand panels) and whether autocorrelation has a causal impact on algorithmic trading activity (right hand panels). Results are presented separately for each currency pair and for each of the following measures of algorithmic trading activity: Overall AT participation (*VAT*), AT taking participation (*VCT*), AT making participation (*VCm*), Relative Computer Taking (*OFCT*), and the log of AT trade correlation ( $\log(R)$ ). All results are based on 5-minute data covering the full sample period from 2003 to 2007. The first seven rows in each panel presents the results from three different Granger causality tests, based on the reduced form VAR in equation (??). In particular, the first two rows in each panel reports the coefficient estimate and standard error of the first lag-coefficient for the causing variable. Rows three to five report the sum of the lag-coefficients for the causing variable, along with the corresponding Wald  $\chi^2$ -statistic and p-value for the null hypothesis that this sum is equal to zero, respectively. The sixth and seventh row report the Wald  $\chi^2$ -statistic and p-value for the null hypothesis that the coefficients on all lags of the causing variable are jointly equal to zero. The following row, labeled Contemp. coeff., presents the point estimate of the contemporaneous impact of the causing variable in the structural VAR in equation (??), based on the heteroskedasticity identification scheme described in the main text; the Newey-West standard error is presented below in parentheses. The last five rows in each panel show, respectively, the total number of observations available for estimation in the full sample, the number of observations available in each of the two sub samples used in the heteroskedasticity identification scheme, and number of different days that are included in each of these two sub samples. The \*\*\*, \*\*, and \* represent a statistically significant deviation from zero at the 1, 5, and 10 percent level, respectively.

	USD/EUR	JPY/USD	JPY/EUR	USD/EUR	JPY/USD	JPY/EUR
Tests of <i>VAT</i> Causing Autocorrelation						
Sum of coeffs. on <i>VAT</i> lags	-0.03904***	-0.02562***	-0.01539***	Sum of coeffs. on <i>ac</i> lags	-0.00689*	-0.01029*
$\chi^2$ (Sum = 0)	33.7507	25.2649	14.9665	$\chi^2_1$ (Sum = 0)	3.0155	3.4939
p-value	0.0000	0.0000	0.0001	p-value	0.0825	0.0616
$\chi^2_{20}$ (All coeffs. on <i>VAT</i> lags = 0)	43.0591***	45.5455***	15.1881***	$\chi^2_{20}$ (All coeffs. on <i>ac</i> lags = 0)	12.8762**	4.0260
p-value	0.0000	0.0000	0.0043	p-value	0.0119	0.4025
Contemp. coeff.	-0.04770***	-0.02377***	-0.03274***	Contemp. coeff.	0.00330*	0.00100
(std. err.)	(0.00703)	(0.00622)	(0.00927)	(std. err.)	(0.00188)	(0.00427)
No. of obs.	102432	102427	102113	No. of obs.	102432	102427
No. of obs. in 1st sub sample	51264	51260	51019	No. of obs. in 1st sub sample	51264	51260
No. of obs. in 2nd sub sample	51168	51167	51094	No. of obs. in 2nd sub sample	51168	51167
No. of unique days in 1st sub sample	534	534	534	No. of unique days in 1st sub sample	534	534
No. of unique days in 2nd sub sample	533	533	533	No. of unique days in 2nd sub sample	533	533
Tests of <i>VCT</i> Causing Autocorrelation						
Sum of coeffs. on <i>VCT</i> lags	-0.04325***	-0.02187***	-0.01069**	Sum of coeffs. on <i>ac</i> lags	-0.00019	-0.00123
$\chi^2$ (Sum = 0)	23.0192	11.8863	6.1195	$\chi^2_1$ (Sum = 0)	0.0033	0.0688
p-value	0.0000	0.0006	0.0134	p-value	0.9545	0.7931
$\chi^2_{20}$ (All coeffs. on <i>VCT</i> lags = 0)	24.7961***	20.8758***	6.6594	$\chi^2_{20}$ (All coeffs. on <i>ac</i> lags = 0)	2.6777	1.7370
p-value	0.0001	0.0003	0.1550	p-value	0.6131	0.7840
Contemp. coeff.	-0.02630***	-0.00557	-0.00255	Contemp. coeff.	0.00138	-0.00343
(std. err.)	(0.00752)	(0.00559)	(0.00567)	(std. err.)	(0.00105)	(0.00247)
No. of obs.	102432	102427	102113	No. of obs.	102432	102427
No. of obs. in 1st sub sample	51264	51260	51019	No. of obs. in 1st sub sample	51264	51260
No. of obs. in 2nd sub sample	51168	51167	51094	No. of obs. in 2nd sub sample	51168	51167
No. of unique days in 1st sub sample	534	534	534	No. of unique days in 1st sub sample	534	534
No. of unique days in 2nd sub sample	533	533	533	No. of unique days in 2nd sub sample	533	533

Table 5: High-frequency autocorrelation and algorithmic trading (cont.).

	USD/EUR	JPY/USD	JPY/EUR	USD/EUR	JPY/USD	JPY/EUR
Tests of $VCm$ Causing Autocorrelation						
Sum of coeffs. on $VCm$ lags	-0.04692***	-0.03620***	-0.02146***	-0.00917***	-0.01358***	-0.02662***
$\chi^2_1$ (Sum = 0)	29.7277	28.8421	17.8840	9.3691	10.1188	20.1782
p-value	0.0000	0.0000	0.0000	0.0022	0.0015	0.0000
$\chi^2_{20}$ (All coeffs. on $VCm$ lags = 0)	44.1685***	38.8638***	18.9016***	33.8031***	11.0068**	20.2980***
p-value	0.0000	0.0000	0.0008	0.0000	0.0265	0.0004
Contemp. coeff.	-0.07073***	-0.03907***	-0.02409***	0.00156	0.00268	0.03122***
(std. err.)	(0.00902)	(0.00711)	(0.00622)	(0.00137)	(0.00269)	(0.00571)
No. of obs.	102432	102427	102113	102432	102427	102113
No. of obs. in 1st sub sample	51264	51260	51019	51264	51260	51019
No. of obs. in 2nd sub sample	51168	51167	51094	51168	51167	51094
No. of unique days in 1st sub sample	534	534	534	534	534	534
No. of unique days in 2nd sub sample	533	533	533	533	533	533
Tests of $VCm$ Causing Autocorrelation						
Sum of coeffs. on $VCm$ lags	-0.00481	-0.00154	0.00013	-0.02571**	-0.01388	-0.01343
$\chi^2_1$ (Sum = 0)	2.3004	0.2637	0.0021	5.0050	1.2702	1.0988
p-value	0.1293	0.6076	0.9638	0.0253	0.2597	0.2945
$\chi^2_{20}$ (All coeffs. on $VCm$ lags = 0)	2.8266	1.8544	1.3831	8.5794*	4.7680	2.7475
p-value	0.5872	0.7625	0.8471	0.0725	0.3119	0.6009
Contemp. coeff.	0.00434	-0.00631	-0.00124	-0.00958	0.03736**	0.01716
(std. err.)	(0.00455)	(0.00463)	(0.00464)	(0.01482)	(0.01857)	(0.02067)
No. of obs.	102250	101992	100815	102250	101992	100815
No. of obs. in 1st sub sample	51095	50871	49935	51095	50871	49935
No. of obs. in 2nd sub sample	51155	51121	50880	51155	51121	50880
No. of unique days in 1st sub sample	534	534	534	534	534	534
No. of unique days in 2nd sub sample	533	533	533	533	533	533
Tests of $OFcT$ Causing Autocorrelation						
Sum of coeffs. on $OFcT$ lags	-0.00481	-0.00154	0.00013	-0.02571**	-0.01388	-0.01343
$\chi^2_1$ (Sum = 0)	2.3004	0.2637	0.0021	5.0050	1.2702	1.0988
p-value	0.1293	0.6076	0.9638	0.0253	0.2597	0.2945
$\chi^2_{20}$ (All coeffs. on $OFcT$ lags = 0)	2.8266	1.8544	1.3831	8.5794*	4.7680	2.7475
p-value	0.5872	0.7625	0.8471	0.0725	0.3119	0.6009
Contemp. coeff.	0.00434	-0.00631	-0.00124	-0.00958	0.03736**	0.01716
(std. err.)	(0.00455)	(0.00463)	(0.00464)	(0.01482)	(0.01857)	(0.02067)
No. of obs.	102250	101992	100815	102250	101992	100815
No. of obs. in 1st sub sample	51095	50871	49935	51095	50871	49935
No. of obs. in 2nd sub sample	51155	51121	50880	51155	51121	50880
No. of unique days in 1st sub sample	534	534	534	534	534	534
No. of unique days in 2nd sub sample	533	533	533	533	533	533
Tests of $\log(R)$ Causing Autocorrelation						
Coeff. on first $\log(R)$ lag	-0.12498**	-0.03268	0.02066	0.00011	-0.00008	-0.00123***
(std. err.)	(0.06255)	(0.05501)	(0.05456)	(0.00032)	(0.00039)	(0.00047)
Sum of coeffs. on $\log(R)$ lags	-0.22045*	-0.10045	0.19150*	-0.00011	-0.00056	-0.00172*
$\chi^2_1$ (Sum = 0)	3.6279	0.9306	3.2797	0.0337	0.5365	3.5320
p-value	0.0568	0.3347	0.0701	0.8543	0.4639	0.0602
$\chi^2_{20}$ (All coeffs. on $\log(R)$ lags = 0)	5.6540	1.8232	3.8823	3.4404	2.3862	9.7551**
p-value	0.2265	0.7682	0.4222	0.4870	0.6651	0.0448
Contemp. coeff.	0.32317	0.57153	-0.00193	-0.00349**	-0.00571*	-0.00120
(std. err.)	(0.26693)	(0.41419)	(0.15163)	(0.00136)	(0.00297)	(0.13197)
No. of obs.	50166	46115	38653	50166	46115	38653
No. of obs. in 1st sub sample	5860	5258	5236	5860	5258	5236
No. of obs. in 2nd sub sample	44306	40857	33417	44306	40857	33417
No. of unique days in 1st sub sample	255	329	335	255	329	335
No. of unique days in 2nd sub sample	533	533	533	533	533	533



Table 6: Triangular arbitrage, AT activity and AT trade correlation.

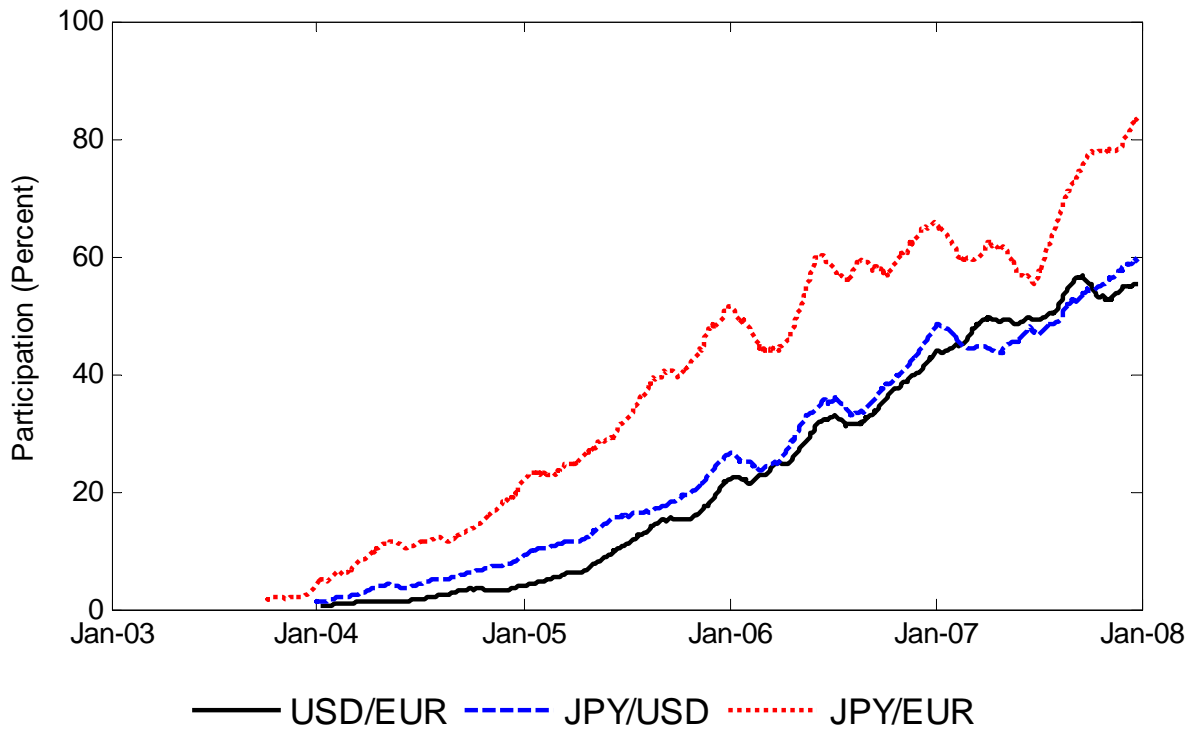
We report tests of whether algorithmic trading activity, measured as Relative Computer Taking (*OFCT*), has a causal impact on triangular arbitrage (top left hand panel), whether triangular arbitrage has a causal impact on *OFCT* (top right hand panel), whether the log of AT trade correlation ( $\log(R)$ ) has a causal impact on triangular arbitrage (bottom left hand panel), and whether triangular arbitrage has a causal impact on  $\log(R)$  (bottom right hand panel). All results are based on 1-minute data covering the full sample period from 2003 to 2007. The first seven rows in each panel presents the results from three different Granger causality tests, based on the reduced form VAR in equation (??). In particular, the first two rows in each panel reports the coefficient estimate and standard error of the first lag-coefficient for the causing variable. Rows three to five report the sum of the lag-coefficients for the causing variable, along with the corresponding Wald  $\chi^2$ -statistic and p-value for the null hypothesis that this sum is equal to zero, respectively. The sixth and seventh row report the Wald  $\chi^2$ -statistic and p-value for the null hypothesis that the coefficients on all lags of the causing variable are jointly equal to zero. The following row, labeled Contemp. coeff., presents the point estimate of the contemporaneous impact of the causing variable in the structural VAR in equation (??), based on the heteroskedasticity identification scheme described in the main text; the Newey-West standard error is presented below in parentheses. The last five rows in each panel show, respectively, the total number of observations available for estimation in the full sample, the number of observations available in each of the two sub samples used in the heteroskedasticity identification scheme, and number of different days that are included in each of these two sub samples. The \*\*\*, \*\*, and \* represent a statistically significant deviation from zero at the 1, 5, and 10 percent level, respectively.

Tests of <i>OFCT</i> Causing Triangular Arbitrage		Tests of Triangular Arbitrage Causing <i>OFCT</i>	
Coeff. on first <i>OFCT</i> lag	-0.0032***	Coeff. on first <i>arb</i> lag	0.0274**
(std. err.)	(0.0005)	(std. err.)	(0.0137)
Sum of coeffs. on <i>OFCT</i> lags	-0.0064***	Sum of coeffs. on <i>arb</i> lags	0.0376***
$\chi^2$ (Sum = 0)	10.9296	$\chi^2$ (Sum = 0)	33.9744
p-value	0.0009	p-value	0.0000
$\chi^2_0$ (All coeffs. on <i>OFCT</i> lags = 0)	69.2054***	$\chi^2_0$ (All coeffs. on <i>arb</i> lags = 0)	63.9869***
p-value	0.0000	p-value	0.0000
Contemp. coeff.	-0.0224**	Contemp. coeff.	0.6315***
(std. err.)	(0.0111)	(std. err.)	(0.2329)
No. of obs.	147114	No. of obs.	147114
No. of obs. in 1st sub sample	8530	No. of obs. in 1st sub sample	8530
No. of obs. in 2nd sub sample	138584	No. of obs. in 2nd sub sample	138584
No. of unique days in 1st sub sample	184	No. of unique days in 1st sub sample	184
No. of unique days in 2nd sub sample	533	No. of unique days in 2nd sub sample	533
Tests of $\log(R)$ Causing Triangular Arbitrage		Tests of Triangular Arbitrage Causing $\log(R)$	
Coeff. on first $\log(R)$ lag	0.0201**	Coeff. on first <i>arb</i> lag	0.0031***
(std. err.)	(0.0084)	(std. err.)	(0.0008)
Sum of coeffs. on $\log(R)$ lags	-0.0286	Sum of coeffs. on <i>arb</i> lags	0.0009**
$\chi^2$ (Sum = 0)	0.6805	$\chi^2$ (Sum = 0)	5.5112
p-value	0.4094	p-value	0.0189
$\chi^2_0$ (All coeffs. on $\log(R)$ lags = 0)	25.1592	$\chi^2_0$ (All coeffs. on <i>arb</i> lags = 0)	45.2746***
p-value	0.1954	p-value	0.0010
Contemp. coeff.	-0.8748**	Contemp. coeff.	0.1102***
(std. err.)	(0.4446)	(std. err.)	(0.0407)
No. of obs.	147114	No. of obs.	147114
No. of obs. in 1st sub sample	8530	No. of obs. in 1st sub sample	8530
No. of obs. in 2nd sub sample	138584	No. of obs. in 2nd sub sample	138584
No. of unique days in 1st sub sample	184	No. of unique days in 1st sub sample	184
No. of unique days in 2nd sub sample	533	No. of unique days in 2nd sub sample	533

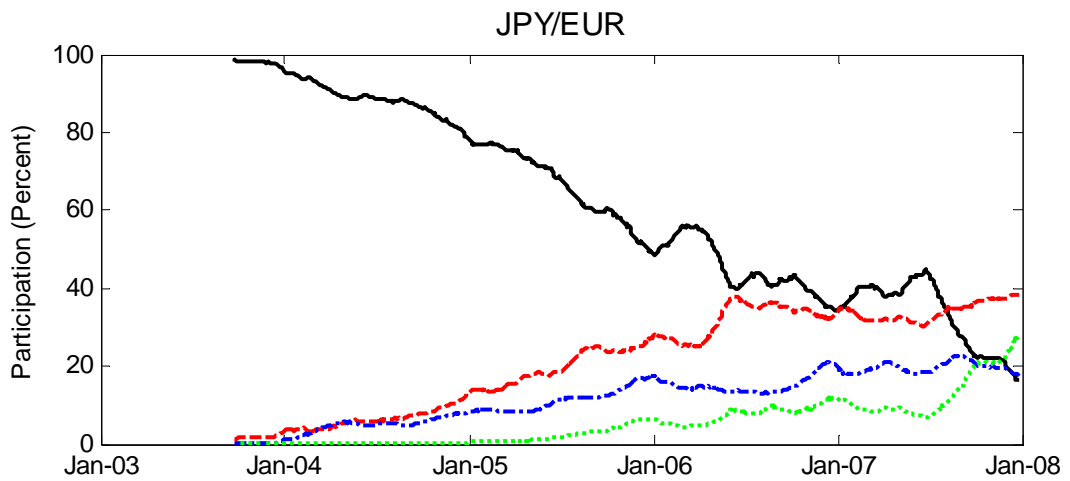
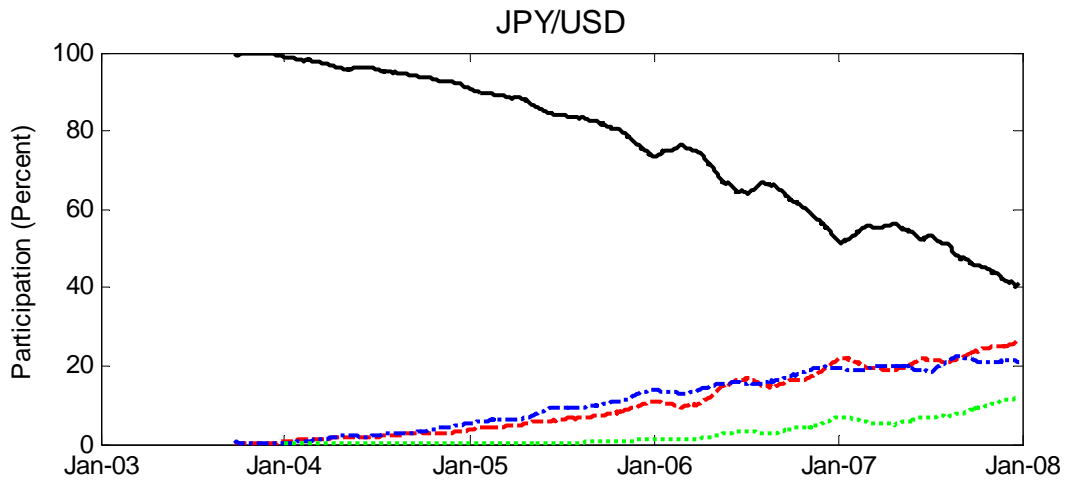
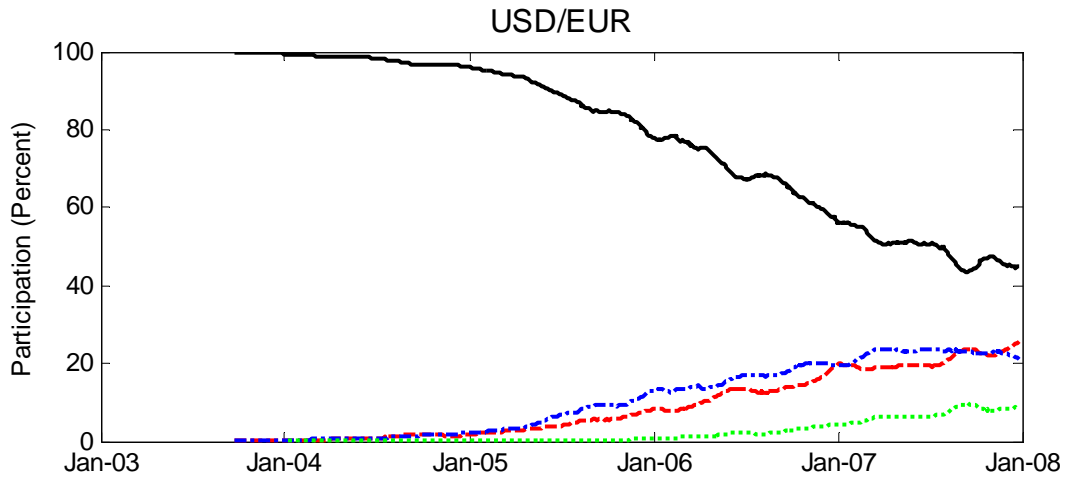
Table 7: High-frequency autocorrelation, AT activity and AT trade correlation.

We report tests of whether algorithmic trading activity, measured as AT making participation ( $VCm$ ), has a causal impact on autocorrelation (top left hand panel), whether autocorrelation has a causal impact on  $VCm$  (top right hand panel), whether the log of AT trade correlation ( $\log(R)$ ) has a causal impact on autocorrelation (bottom left hand panel), and whether autocorrelation has a causal impact on  $\log(R)$  (bottom right hand panel). All results are based on 1-minute data covering the full sample period from 2003 to 2007. The first seven rows in each panel presents the results from three different Granger causality tests, based on the reduced form VAR in equation (??), estimated separately for each currency pair. In particular, the first two rows in each panel reports the coefficient estimate and standard error of the first lag-coefficient for the causing variable. Rows three to five report the sum of the lag-coefficients for the causing variable, along with the corresponding Wald  $\chi^2$ -statistic and p-value for the null hypothesis that the coefficients on all lags of the causing variable are jointly equal to zero. The sixth and seventh row report the Wald  $\chi^2$ -statistic and p-value for the null hypothesis that the contemporaneous impact of the causing variable in the structural VAR in equation (??), based on the heteroskedasticity identification scheme described in the main text; the Newey-West standard error is presented below in parentheses. The last five rows in each panel show, respectively, the total number of observations available for estimation in the full sample, the number of different days that are included in each of the two sub samples used in the heteroskedasticity identification scheme, and number of different days that are included in each of these two sub samples. The \*\*\*, \*\*, and \* represent a statistically significant deviation from zero at the 1, 5, and 10 percent level, respectively.

	USD/EUR	JPY/USD	JPY/EUR	USD/EUR	JPY/USD	JPY/EUR
Tests of $VCm$ Causing Autocorrelation						
Coeff. on first $VCm$ lag	-0.03137***	-0.00743	-0.00754	-0.09039	-0.02261	0.02479
(std. err.)	(0.00713)	(0.00539)	(0.00478)	(0.06291)	(0.05534)	(0.05461)
Sum of coeffs. on $VCm$ lags	-0.04854***	-0.02346***	-0.00877	-0.15189	-0.06853	0.19624*
$\chi^2$ (Sum = 0)	23.7165	7.8819	1.4877	1.6974	0.4280	3.4397
p-value	0.0000	0.0050	0.2226	0.1926	0.5130	0.0636
$\chi^2_0$ (All coeffs. on $VCm$ lags = 0)	36.9543***	8.3181*	4.6417	3.4351	1.2139	3.9416
p-value	0.0000	0.0806	0.3261	0.4878	0.8758	0.4140
Contemp. coeff.	-0.07867***	-0.06572***	-0.02133	0.38640	0.67458*	0.13013
(std. err.)	(0.02839)	(0.02270)	(0.23757)	(0.26970)	(0.40989)	(12.09710)
No. of obs.	50166	46115	38653	50166	46115	38653
No. of obs. in 1st sub sample	5860	5258	5236	5860	5258	5236
No. of obs. in 2nd sub sample	44306	40857	33417	44306	40857	33417
No. of unique days in 1st sub sample	255	329	335	255	329	335
No. of unique days in 2nd sub sample	533	533	533	533	533	533
Tests of $\log(R)$ Causing Autocorrelation						
Coeff. on first $\log(R)$ lag	-0.01457***	-0.00810**	-0.00171	0.00017	-0.00006	-0.00123***
(std. err.)	(0.00281)	(0.00406)	(0.00542)	(0.00032)	(0.00039)	(0.00047)
Sum of coeffs. on $\log(R)$ lags	-0.01581***	-0.02371***	-0.01799*	0.00014	-0.00047	-0.00172*
$\chi^2$ (Sum = 0)	8.5663	9.1399	2.9671	0.0482	0.3888	3.5365
p-value	0.0034	0.0025	0.0850	0.8261	0.5330	0.0600
$\chi^2_0$ (All coeffs. on $\log(R)$ lags = 0)	27.6599***	11.6210**	10.2542**	3.3473	2.3149	9.6986**
p-value	0.0000	0.0204	0.0364	0.5015	0.6781	0.0458
Contemp. coeff.	0.00823	0.03076*	0.03639	-0.00338**	-0.00582*	-0.00157
(std. err.)	(0.01069)	(0.01637)	(0.02563)	(0.00137)	(0.00299)	(0.11278)
No. of obs.	50166	46115	38653	50166	46115	38653
No. of obs. in 1st sub sample	5860	5258	5236	5860	5258	5236
No. of obs. in 2nd sub sample	44306	40857	33417	44306	40857	33417
No. of unique days in 1st sub sample	255	329	335	255	329	335
No. of unique days in 2nd sub sample	533	533	533	533	533	533

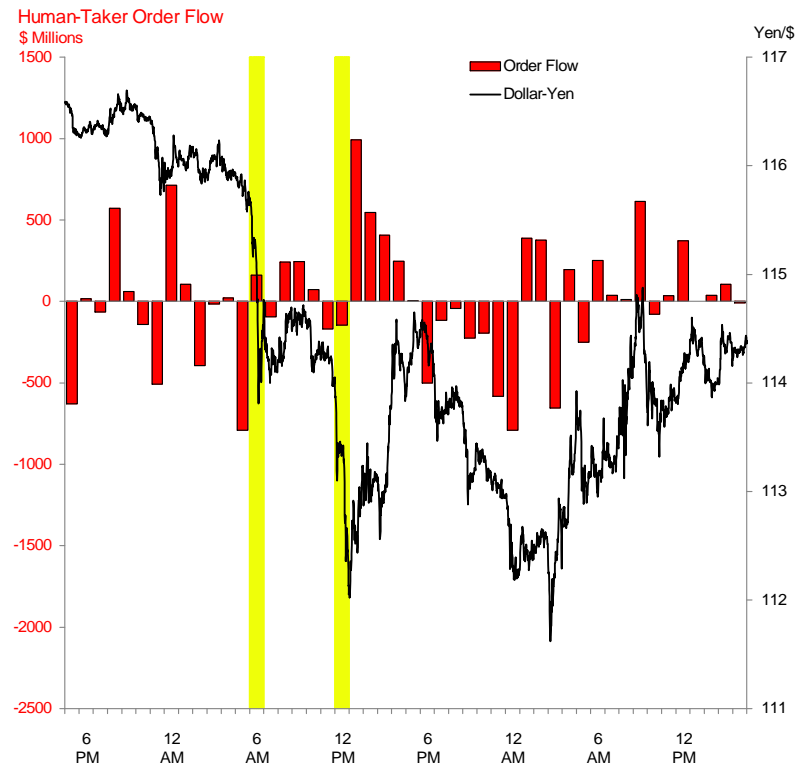
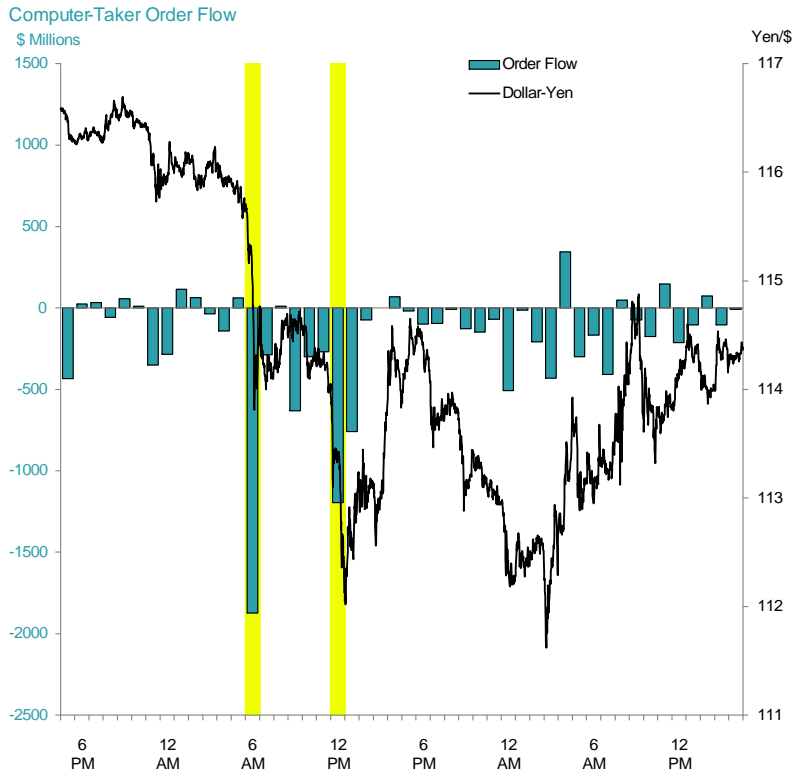


**Figure 1: 50-day moving averages of participation rates of algorithmic traders**

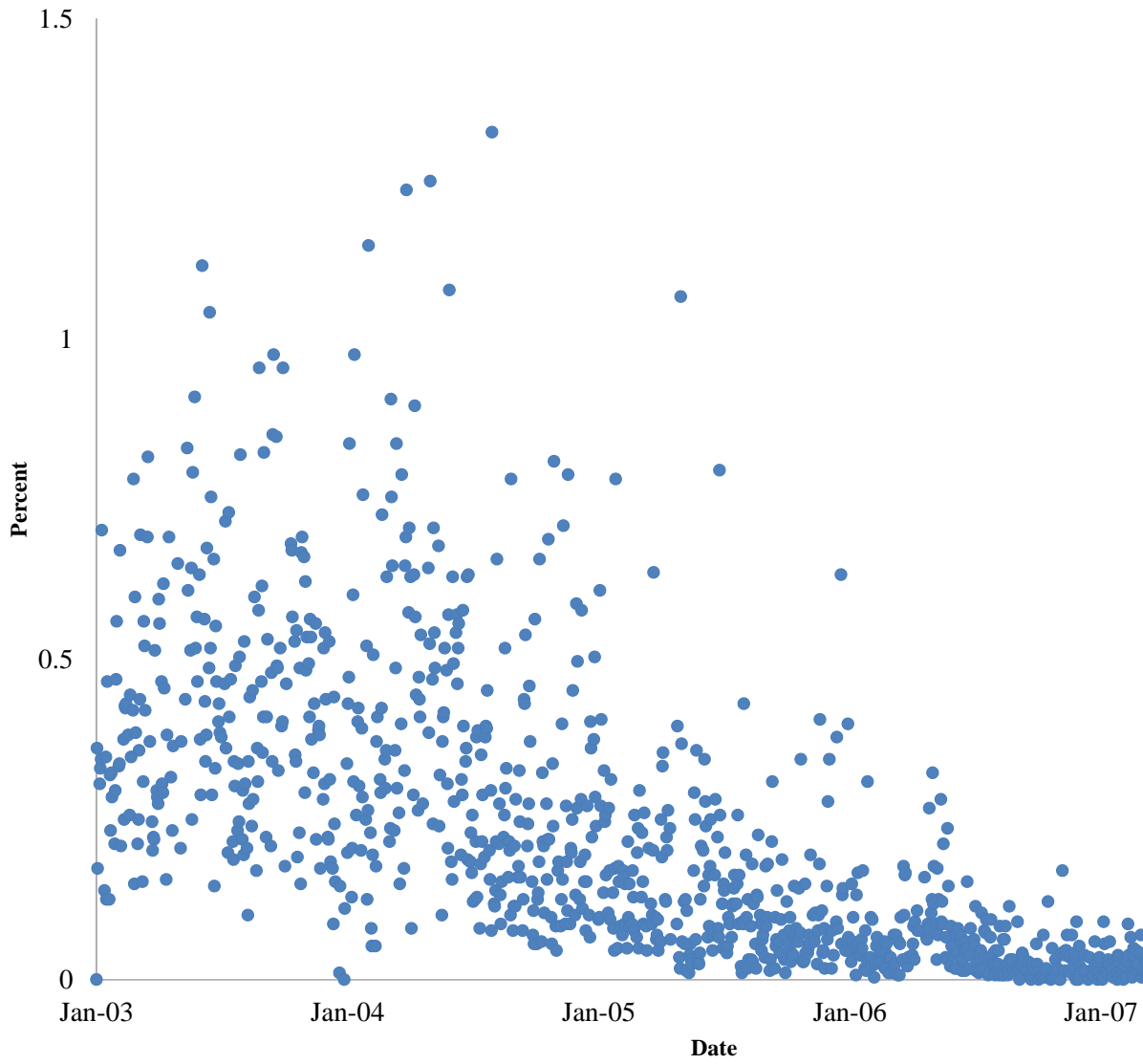


— H-Maker/H-Taker                      - - - C-Maker/H-Taker  
 - - - H-Maker/C-Taker                      ····· C-Maker/C-Taker

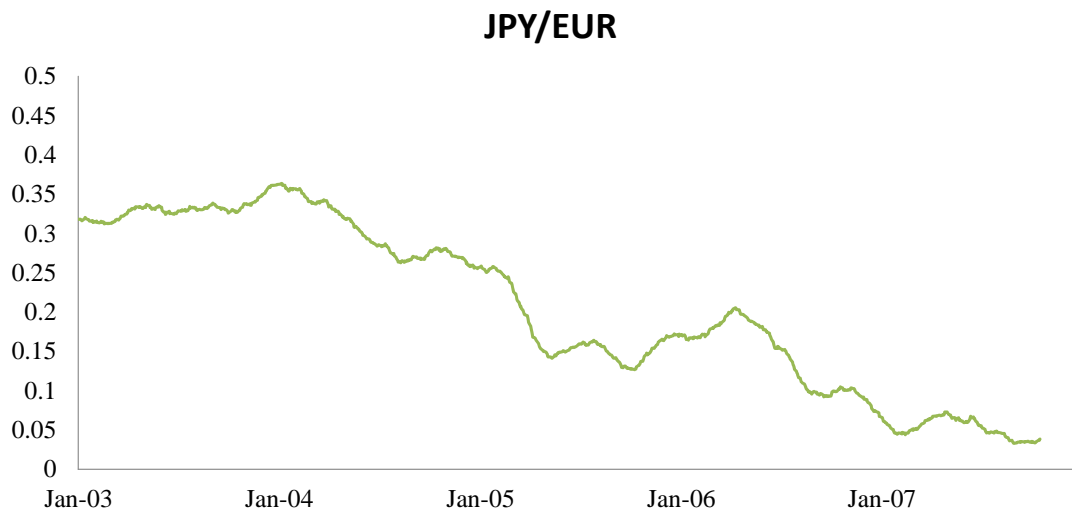
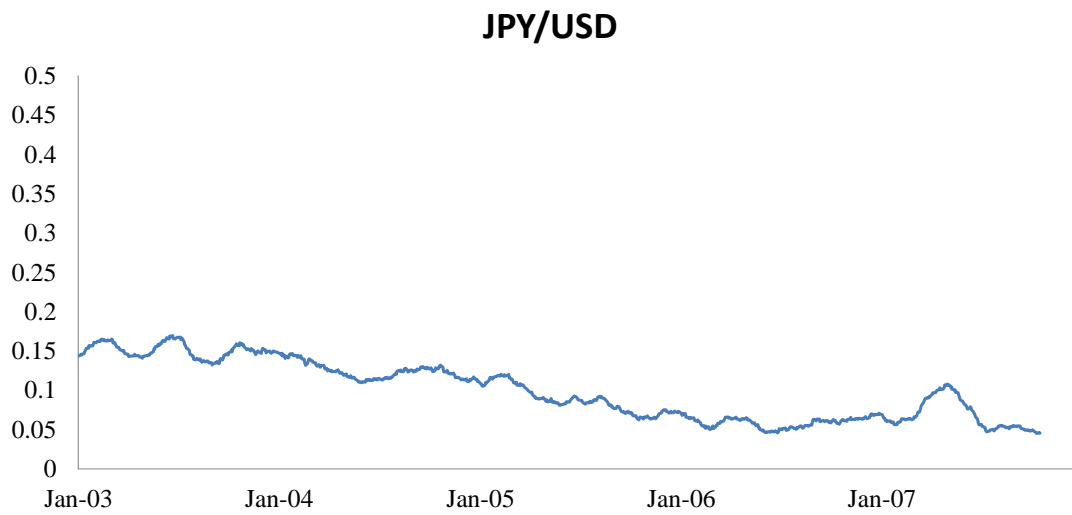
**Figure 2: 50-day moving averages of participation rates broken down into four maker-taker pairs**



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



**Figure 4: Percent of seconds with a triangular arbitrage opportunity with a profit greater than 1 basis point within the busiest trading hours, 3:00 to 11:00 am ET, of the day**



**Figure 5: 50-day moving averages of absolute value of 5-second return serial autocorrelation estimated each day using observations from 3:00 am ET to 11:00 am ET.**