



HAL
open science

A new wavelet-based ultra-high-frequency analysis of triangular currency arbitrage

Nikola Gradojevic, Deniz Erdemlioglu, Ramazan Gençay

► **To cite this version:**

Nikola Gradojevic, Deniz Erdemlioglu, Ramazan Gençay. A new wavelet-based ultra-high-frequency analysis of triangular currency arbitrage. *Economic Modelling*, 2020, 85, pp.57-73. 10.1016/j.econmod.2019.05.006 . hal-02512423

HAL Id: hal-02512423

<https://hal.science/hal-02512423v1>

Submitted on 21 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

A new wavelet-based ultra-high-frequency analysis of triangular currency arbitrage

Nikola Gradojevic* Ramazan Gençay† Deniz Erdemlioglu‡

Abstract

We develop a new framework to characterize the dynamics of triangular (three-point) arbitrage in electronic foreign exchange markets. To examine the properties of arbitrage, we propose a wavelet-based regression approach that is robust to estimation errors, measurement bias and persistence. Relying on this wavelet-based (denoising) inference, we consider various liquidity and market risk indicators to predict arbitrage in a unique ultra-high-frequency exchange rate data set. We find strong empirical evidence that limit order book, realized volatility and cross-correlations help forecast triangular arbitrage profits. The estimates are statistically significant and relevant for investors such that on average 80–100 arbitrage opportunities exist with a short duration (100–500 milliseconds) on a daily basis. Our analysis also reveals that triangular arbitrage opportunities are counter-cyclical at ultra-high-frequency levels: arbitrage returns tend to increase (decrease) in periods when volatility risk and correlations are relatively low (high). We show that liquidity-driven microstructure measures, however, appear to be more powerful in exploiting arbitrage profits when compared to market-driven factors.

Keywords: Foreign Exchange Markets; Triangular Arbitrage; Limit Order Book; Wavelets

JEL No: G15, G17, F31

*Corresponding author. University of Guelph, College of Business and Economics, Department of Economics and Finance, 50 Stone Road East, Guelph, Ontario, N1G 2W1, Canada; IÉSEG School of Management (LEM-CNRS), France; Phone: Tel.: (+1) 519-824-4120. Email: ngradoje@uoguelph.ca. The authors are grateful to the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada for research support. We also thank Thierry Foucault, Lawrence Harris, Terrence Hendershott, Angelo Ranaldo, Ilias Tsiakas, Ege Yazgan and seminar participants at the 2nd Rimini Time Series workshop (2013) and the Rimini Conference in Economics and Finance (2014) for their helpful comments and suggestions on earlier versions of the paper.

†Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada. Email: gencay@sfu.ca. Ramazan Gençay passed away during this research.

‡IÉSEG School of Management (LEM-CNRS), Department of Audit, Finance and Control. 3 Rue de la digue, 59000 Lille, France. Email: d.erdemlioglu@ieseg.fr.

1. Introduction

“Prediction is very difficult, especially if it’s about the future.”

(Neils Bohr)

Arbitrage in finance is broadly viewed as an attractive investment strategy because it offers riskless return with no cost. If the arbitrage opportunity exists as a result of certain market anomalies or inefficiencies, a trader can simultaneously buy and sell a financial asset to profit from price differentials (Wu and Chen, 2008). In reality, however, there is no clear consensus whether such arbitrage opportunities exist and how to measure, explain and test for the predictability of arbitrage returns. Compared to stocks and bonds, the challenge intensifies even further for currency markets where the arbitrage has to be formed by three assets, each having different statistical regularities and trading process. Researchers and engineers study this unique form of arbitrage as *triangular (three-point) arbitrage*.

This paper develops a new methodological and empirical framework to characterize the dynamics of triangular currency arbitrage at ultra-high-frequency trading scales. Our prediction technique relies on a wavelet filtration approach that controls for estimation errors and any potential inference biases. Using an exclusive ultra-high-frequency data from an electronic foreign exchange (FX) trading platform, we consider wavelet-based regressions to predict one-step ahead triangular arbitrage. Extending this setup, we further seek to identify high-frequency determinants of triangular arbitrage returns through market-wide information (i.e., volatility, correlation) and trading-implied factors (i.e., bid-ask spread, center of gravity). Our analysis provides relevant financial implications for FX market efficiency (Giannellis and Papadopoulos, 2009), international market microstructure mechanisms and their role in FX rate formation (Gençay and Gradojevic, 2013).

More specifically, we position our study in the recent literature on triangular arbitrage in several ways. For instance, Darolles *et al.* (2017) propose a model that exploits triangular structure to identify liquidity frictions and volume characteristics. Examining the role of scheduled news events in this context, Mueller *et al.* (2017) find evidence of significant excess returns occurring as compensation for monetary policy uncertainty on days with FOMC announcements. In this direction, Piccotti (2018) empirically links price jumps to triangular arbitrage and finds that illiquidity cost is responsible for limiting arbitrage, rather than the jump movement itself. Through the lens of market microstructure, Cui *et al.* (2018) introduce a new computational framework that augments the triangular arbitrage scheme to a k -currency arbitrage, and show how to detect multiple forms of arbitrage opportunities using high frequency data on the G-10 exchange rates. From a macroeconomic perspective, Aloosh and Bekaert (2019) recently develop a factor model to capture the

dynamics of the G-10 currencies by introducing the notion of *currency baskets*.

Our paper extends this new line of research methodologically and empirically.¹ Darolles *et al.* (2017) deal with the information (flow) problem and uncover the non-negligible role of the strategic behavior of long-term investors in lowering daily price efficiencies. In contrast, our modeling focus is mainly on ultra-high-frequency trading and thus the implications of our results are relatively more essential for short-term traders, who quite often prefer to rebalance positions in short periods of time (e.g., seconds). Mueller *et al.* (2017) work with currency portfolio returns, but assume that there is no violation of triangular arbitrage. Rather, we directly test for the presence of triangular arbitrage. The analysis of Aloosh and Bekaert (2019) aims at controlling for the triangular arbitrage whereas, instead, we exploit and characterize arbitrage opportunities. The methodological setup of Piccotti (2018) reveals how jump type events create FX mispricing or the so-called triangular pricing error. Although Piccotti (2018) utilizes maximum quote revision in a 5-minute period, it is extremely hard, if not impossible, to detect true triangular arbitrage (with low likelihood of false discovery) at such minute-based scales. We use data at the ultra-high-frequency and adopt realized volatility measures, instead of price jumps.

The central argument in our research is that triangular arbitrage operates in the millisecond trading environment that accommodates the actions of high-frequency traders. We therefore investigate whether the analysis of the shape of the limit order book can benefit high-frequency trading and thus affect the FX market quality as measured by the average bid-ask spread. The focus is on the notion that higher transaction costs reduce triangular arbitrage opportunities and limit the role of high-frequency traders. However, lower average trading costs increase the size and likelihood of triangular arbitrage profits, which may further improve market quality. In addition, we attempt to understand how the average FX market volatility in the three exchange rates from the triangular parity condition influences arbitrage profits. Specifically, we conjecture that in the presence of lower average volatility, sudden market shocks can have larger impacts and produce shifts in one or more exchange rate quotes, which may result in an arbitrage opportunity. Hence, the key message of our paper is that FX traders may profit from both the pace of information gathering (e.g., information related to limit book indicators, and average variance and correlation measures) and high-frequency order placement that is based on this information. In essence, an improved FX market quality attracts high-frequency traders who identify and exhaust triangular arbitrage opportunities. Meanwhile, human FX traders are relatively slow and thus less likely to benefit from the reduction in the average spread and average variance (correlations).

After conducting an extensive search for bid and ask price misalignment from the triangular

¹In Section 2, we review the related literature thoroughly and aim at discussing the contributions of our paper to extant literature more closely.

parity condition, we find on average about 80-100 such instances in the data on a daily basis. These profitable deviations are short-lived with durations between 100 and 500 milliseconds. The average triangular arbitrage strategy return is in the range from 0.5-0.75 basis points (bps). In line with our conjecture, the findings also show that both volatility and correlation measures that we employ are informative for explaining and predicting triangular arbitrage returns. In particular, higher average volatility reduces arbitrage returns both contemporaneously and in a predictive setting. We conclude that dealers in the interbank market become more “watchful” about updating their quotes when high FX volatility is observed, thus, improving the FX market efficiency. Furthermore, we find that when average correlations across the three FX rates are low, triangular arbitrage returns are expected to be higher. In other words, triangular arbitrage opportunities are more frequent in the times when the average degree of interaction among the FX rates is low, which introduces potential impediments to synchronous adjustments of exchange rates to market shocks.

In addition to the empirical evidence supporting predictability of risk-induced arbitrage returns, we find that limit order book indicators are informative and useful. Measures, such as the average inside bid-ask spread and the quantity-weighted average bid-ask spread (reflecting the shape of the limit order book) are statistically significant in explaining and predicting arbitrage profits. In general, we find that tighter average spread measures increase the profitability and likelihood of arbitrage trades. Intuitively, spreads are narrower during liquid periods when the volume of high-frequency trading is large and potential price misalignments are more likely to occur. Finally, we uncover a strong level of persistence in all predictors of arbitrage returns and, when we compensate for any potential estimation biases coming from such processes, our predictors exhibit stronger predictive power.

We investigate predictive relationships and time series properties of predictors of arbitrage returns at the highest available frequency level. Our unique high-frequency data set includes ten layers of tick quotes (at the 100 millisecond precision) on the bid and ask sides of the limit order book. This offers an insight into the depth of the limit order book, which includes levels of unrealized currency order flows. The data are taken from Electronic Broking Services (EBS), the major interdealer platform for spot FX trading. To the best of our knowledge, this is the most comprehensive data set currently available for research in high-frequency international finance.² Considering the vast amount of data and considerable computational requirements, we focus on

²Kozhan and Salmon (2012) use Reuters electronic FX trading system at the 1/100th of a second resolution, but do not present the number of layers in their limit order book. The data in equity markets are already available at the nanosecond frequency. EBS has a plan to reduce data frequency to 50 milliseconds. The reason why interbank FX market is different is because, unlike in equity markets, there are different trader types in the interbank FX market including human traders, automated traders and slower high-frequency traders. The market has no interest to provide an ideal environment just for high-frequency traders. Section 6 will provide more insights into this issue.

three major currency pairs (EUR-JPY, EUR-USD and USD-JPY) over several time periods in 2010 and 2011.

Our empirical framework offers several viable implications for investment analysis and FX trading.³ First, we show that triangular arbitrage can be profitable and is to a certain extent predictable in real-time trading at ultra-high-frequencies. Consequently, investment analysts and managers could benefit by incorporating data with (at least) millisecond resolution as part of their trading strategies. Second, when designing a high-frequency arbitrage trading strategy, investors should utilize *wavelet-based arbitrage regressions* in order to assess predictability and profitability. This method is noise-free, easy to implement, computationally tractable and it does not suffer from any estimation-induced biases, such as persistence or serial dependence. It is important to stress that FX market microstructure literature routinely neglects the fact that currency order flows (as well as macroeconomic fundamentals) are subject to measurement errors (Evans and Lyons, 2012), which produces biased and inconsistent OLS estimators. Theoretically speaking, our methodology is aimed at removing high-frequency jumps and structural breaks and what is left is the genuine long-memory part of the spectrum.⁴ Third, in contrast to the conventional wisdom, we explicitly show that triangular arbitrage returns are *endogenous*, which implies that opportunities depend on market conditions, business cycles and trading process itself. In this regard, an investment strategy should consider three statistically significant predictors: realized volatility, realized correlations and liquidity. We emphasize that an arbitrage model based on our technique is likely to perform better in normal rather than crisis or distress periods, that is, when volatility and correlations are low, and when liquidity is high.

The remainder of the paper is organized as follows. The next section discusses in depth the related literature. In Section 3, we introduce the process of triangular arbitrage and define the predictors. Section 4 describes the data set. In Section 5, we outline our estimation scheme based on the wavelet approach. Section 6 presents the empirical results and Section 7 concludes.

³As we will discuss in Section 2, studies on arbitrage (stock or triangular) often focus on tackling optimization-related problems with limited attention to managerial implications. Few exceptions include the works of Ito *et al.* (2012), Kozhan and Salmon (2012) and Aiba and Hatanoa (2004). Ito *et al.* (2012) and Kozhan and Salmon (2012) conclude that arbitrage opportunities are difficult to exploit due to advances in high-speed algorithmic trading and execution costs. Aiba and Hatanoa (2004) find evidence in favor of triangular arbitrage and show that triangular arbitrage interactions (among FX rates) are driven by the sharp peaks and tails of the return distributions.

⁴More broadly, our theoretical (modeling) framework draws from Cenedese *et al.* (2014) and Kozhan and Salmon (2012).

2. Related Literature

On the financial front, the literature on arbitrage dynamics is vast with a particular focus on stocks. In contrast to stock market arbitrage, the nature of triangular currency arbitrage relies on *three* instead of *two* points (as in security markets) and hence requires a different research treatment. Triangular arbitrage strategy is particularly interesting because it is not hindered by taxes, short-selling or funding constraints, and the risk of these strategies is very limited. Hence, standard limits to arbitrage cannot explain why triangular arbitrage opportunities are not eliminated immediately (Pasquariello, 2014). Triangular arbitrage opportunities are similar to other high speed opportunities in that they are (i) frequent, (ii) very short-lived, (iii) more efficiently exploited by machines than by humans, and (iv) they deliver razor blade profits per opportunity (Foucault *et al.*, 2017). Therefore, we examine currency arbitrage and also the drivers of the potential arbitrage profits in high-frequency financial data. While these aforementioned studies explore the implications of arbitrage conditions for designing optimal portfolios, our approach traces *when* arbitrage opportunity arises and how it changes over time at ultra-high-frequency trading scales.⁵ Our results for currency markets show that the use of ultra-high-frequency data from an electronic trading platform conveys information, rather than noise, and allows investors to exploit the profits through trading strategies.

Market volatility and asset correlations have a significant role in financial modeling and prediction (see e.g., Christoffersen and Diebold, 2006, Bali, 2007 and Bajgrowicz *et al.*, 2016). Bali (2007) accounts for nonlinearity and fat-tails in forecasting interest rate volatility. Christoffersen and Diebold (2006) show that forecasting the future direction (i.e., sign forecasting) of financial asset prices heavily depends on the (estimated) volatility level. Similar to the properties of sign forecasting documented in Christoffersen and Diebold (2006), we find evidence that triangular arbitrage profits exhibit time-variation and move with volatility and (cross-rate) correlations. The future direction of the extracted arbitrage spells appears to be the function of realized volatility and correlation. In light of these findings, we can consider a practical implication (a two-step trading strategy) for investors and high-frequency traders. For the periods when volatility is low (and hence in normal times), an investor can test for the volatility timing (as first step) and then

⁵It is worth noting that even triangular arbitrage is detected, FX traders usually face problems, such as the execution risk (i.e., delays in trade execution), “slippage” (i.e., order execution at a worse-than-expected price) and competition from other traders. Kozhan and Tham (2012) stress the importance of execution risk in arbitrage and show simulation evidence that an increase in the number of arbitrageurs reduces arbitrage profits. Fenn *et al.* (2009) argue that there exists high uncertainty of successfully completing all three legs of the arbitrage trade before an unfavorable price movement. Hence, the timing of clearing process may significantly reduce arbitrage profits in practice. In particular, there is a risk that three orders completing triangular arbitrage may not be executed at the same time, mainly due to the market competition for liquidity. In other words, it may be extremely costly to maintain the fastest execution speeds and remain profitable over a prolonged period of time.

exploit the arbitrage return if volatility is below a certain threshold (e.g. α standard deviation from the estimated volatility). Inversely, a spot volatility level exceeding such a threshold is unlikely to generate an arbitrage opportunity. This strategy might also account for correlation patterns: low interest rate currencies often have low correlation and low volatility. From an arbitrage perspective, our results suggest that a successful triangular arbitrage strategy should focus on currencies that have low correlation with others; the lower the cross-correlation, the higher potential for arbitrage returns.

From the methodological standpoint, our paper contributes to the prior research in the frequency-domain analysis, wavelet-based modeling and measurement. For instance, Knif *et al.* (1995) examine the linkages between two European stock markets by utilizing frequency-embedded spectral approach. As Knif *et al.* (1995) show, the multiscale resolution of stock returns and volatility allows to describe market dependence in a tractable way. Haven *et al.* (2012) and Sun and Meinel (2012) use wavelet approach as a filtration (or *denoising*) technique. Through wavelet transform, Haven *et al.* (2012) obtain better out-of-sample forecasts for option prices. While Sun and Meinel (2012) implement the wavelet algorithm (discrete wavelet transform) to high-frequency data from German equity market, Sun *et al.* (2015) focus on U.S. stocks and predict volatility by separating efficient prices from microstructure noise.

Our wavelet-based estimation scheme complements these studies in several important respects. First, extending the model features of Sun and Meinel (2012), we use the maximum overlap discrete wavelet transform (MODWT) to estimate our arbitrage regressions. Second, while Knif *et al.* (1995) provide rather theoretical and simulation-based evidence, we attempt to carry out a comprehensive empirical assessment and aim at providing practical implications for investment analysis. Specifically, we utilize wavelets to denoise the potential estimation errors that typically arise when researchers use high-frequency financial data. In this regard, while the MODWT and GOWDA methods of Sun and Meinel (2012) and Sun *et al.* (2015) control irregularities, such as trend versus noise, our technique performs “robust denoising”, hence further eliminating the potential “error-in-variables” problem in statistical (least square) inference. To the best of our knowledge, we are the first to apply wavelets in regression analysis when ultra-high-frequency data is used. Third, while the option market is of interest for researchers to understand ex-ante (i.e., implied) behavior of investors (as in Haven *et al.*, 2012), arbitrage opportunities are solely linked to real-time trading and thus should be studied through ex-post (i.e., realized) factors. Departing from Haven *et al.* (2012), we thus focus on currency markets and consider ex-post (i.e., volatility and correlation) rather than ex-ante indicators (i.e., implied density forecasts). We empirically show that both volatility and correlation have significant predictive power for arbitrage profits.

More broadly, unlike covered and uncovered interest parity conditions, triangular parity in-

teraction has received very limited attention by the international finance scholars. Perhaps more importantly, the existing literature on this line of research often documents that triangular arbitrage opportunities are scarce and short-lived (Foucault *et al.*, 2017; Fenn *et al.*, 2009; Choi, 2011; Aiba and Hatanoa, 2004). In this regard, Moore and Payne (2011) argue that arbitrage depends on the ability of traders to predict currency order flows. From a time series perspective, Ito *et al.* (2012) use data from 1999 to 2009 and observe an erosion in the number of triangular arbitrage opportunities. We contribute to this literature by developing an alternative estimation technique and providing strong evidence in favor of triangular arbitrage profits. The existence of such arbitrage returns is in line with the recent research on FX market microstructure such that order flows or liquidity measures, for instance, in the EUR/JPY market carry relevant information and impact the EUR/USD and USD/JPY exchange rates (Lyons and Moore, 2009; Daniélsson *et al.*, 2011). Consequently, the deviations from the triangular parity relationship arise and reflect (temporary) market imperfections in at least one of the three markets. We conceptualize and empirically show that investors have enough trading duration to exploit profits from these market imperfections.

It is also important to emphasize that several recent works focused on triangular arbitrage from the *timing* and *profit-making* perspectives (see e.g., Pasquariello, 2014 and Foucault *et al.*, 2017). For instance, Chaboud *et al.* (2014) found that computers can take advantage of short-lived triangular arbitrage opportunities. This paper also explained that hedge funds exploit triangular arbitrage opportunities, often accessing several trading platforms. Furthermore, the Bank for International Settlements—Report submitted by a Study Group established by the Markets Committee (“High-Frequency Trading in the Foreign Exchange Market”)—in 2011 acknowledged the importance of triangular arbitrage and high-frequency trading in the FX market by reporting that it – “operates on high volume but small order sizes, low margins, low latency (with trade execution times measured in milliseconds) and short risk holding periods (typically well under five seconds).” The relevance of triangular arbitrage has also received coverage in the social media. Using the quotes from 2007, Schaumburg (2014) calculated a triangular arbitrage profit of 120.65 EUR per 1 million USD invested, or 0.012 percent.

Our paper attempts to fill the gap in the literature concerning the nature of triangular arbitrage opportunities and their driving forces. This research avenue is novel relative to the previous studies that are primarily concerned with detecting violations of the triangular parity equation, but are silent about measures that may help capture patterns of fluctuations in triangular arbitrage returns. We first relate arbitrage returns to FX risk measures that reflect aggregate movements in volatility and correlation across three exchange rates. Second, we extend this set of measures with electronic limit order book indicators. The key hypothesis that motivates this choice of indicators is that the order book may provide information about future price movements in FX markets. In particular,

such indicators utilize the structure of a limit order book which includes the levels of bid and ask FX rates as well as currency order sizes. The relationship between the shape of the limit order book and arbitrage returns in FX markets has not been covered by previous research, and thus it is explored in the current paper.

3. Main Setup: Triangular Arbitrage and Predictors

In this section, we introduce our main setup. We start by discussing triangular arbitrage strategy. We then proceed with the construction of the realized risk measures and limit order book measures, respectively.

3.1. Triangular Arbitrage Strategy

Triangular parity condition involves three exchange rates $S_{i/j,t}$ ($i \neq j$) that represent FX conversion rates among three currencies at time t (e.g., $i, j \in \{EUR, USD, JPY\}$). When one ignores transaction costs, the triangular parity equation can be written as

$$\frac{S_{EUR/JPY,t}}{S_{USD/JPY,t}} = S_{EUR/USD,t}, \quad (1)$$

where $S_{i/j,t}$ denotes the amount of currency j required to buy one unit of currency i at time t . If Equation 1 does not hold, arbitrage profits may be possible, but the currency conversions have to be executed at the exact FX rates that violated the parity condition. For example, suppose that the initial endowment is one unit of EUR. Then, one can first exchange one EUR for the EUR/JPY amount of JPY. This is followed by a conversion of the JPY to the USD at the USD/JPY exchange rate. Finally, the USD amount is converted to the EUR. In a triangular arbitrage situation, this round trip should produce an amount of EUR that is greater than the initial EUR endowment (i.e., one unit). The other arbitrage route would be to convert one EUR to the USD, then to the JPY and, in the end, to the EUR.

In general, we can start from M units of the EUR currency and follow the first route as

$$(EUR \rightarrow JPY \rightarrow USD \rightarrow EUR), \quad (2)$$

while accounting for the bid-ask spread, the triangular parity condition at time t can be written as

$$\begin{aligned} M \times S_{EUR/JPY,t}^b \times \frac{1}{S_{USD/JPY,t}^a} \times \frac{1}{S_{EUR/USD,t}^a} - M &= 0, \\ S_{EUR/JPY,t}^b \times \frac{1}{S_{USD/JPY,t}^a} \times \frac{1}{S_{EUR/USD,t}^a} - 1 &= 0, \end{aligned} \quad (3)$$

where superscript b denotes the bid quote and superscript a denotes the ask (offer) quote. To provide an illustrative example, we assume that the quotes at time t are given as follows⁶:

- $S_{EUR/USD,t}^a = 1.3911, S_{EUR/USD,t}^b = 1.3909,$
- $S_{EUR/JPY,t}^a = 111.96, S_{EUR/JPY,t}^b = 111.94,$
- $S_{USD/JPY,t}^a = 80.48, S_{USD/JPY,t}^b = 80.47.$

Assuming $M=100$ EUR, the first conversion is to the JPY, by using $S_{EUR/JPY,t}^b = 111.94$. The amount of JPY required to purchase 100 EUR is 11,194 JPY (100×111.94). Next, the amount of 11,194 JPY is converted to the USD as $11,194 / S_{USD/JPY,t}^a$, which produces 139.0905 USD. Finally, we convert back to the EUR by dividing the USD amount by $S_{EUR/USD,t}^a$: $139.0905 / 1.3911 = 99.985$ EUR. In this example, the triangular parity condition does not hold, but the difference is negative ($99.985 - 100 < 0$), which indicates a loss to the arbitrage strategy. The triangular parity condition for the second route (EUR→USD→JPY→EUR) can be expressed as

$$S_{EUR/USD,t}^b \times S_{USD/JPY,t}^b \times \frac{1}{S_{EUR/JPY,t}^a} - 1 = 0. \quad (4)$$

Equations 3 and 4 represent all possible triangular arbitrage parity relationships for this set of exchange rates. If the left-hand side of the two equations is greater than zero at time t , arbitrage profits are possible.

3.2. Realized Market Measures: Triangular Variance and Correlation

This set of predictors is aimed at exploring risk measures that are specific to the FX market. Our measures are adapted to the triangular parity setting and involve three currencies, as opposed to all market exchange rates. Specifically, we are interested in capturing the joint variance and correlation among the EUR-JPY (denoted by “1” for the remainder of the equations in this subsection),

⁶These are the actual high-frequency quotes taken from EBS on November 1, 2010.

EUR-USD (denoted by “2”) and USD-JPY (denoted by “3”) exchange rates. First, we define the “triangular” average return at time $t+1$ as

$$r_{T,t+1} = \frac{1}{3} \sum_{j=1}^3 r_{j,t+1}, \quad (5)$$

where $r_{j,t+1} = \ln(S_{j,t+1}) - \ln(S_{j,t})$ is a standard one-period return from time t to time $t+1$ on the exchange rate S_j ($j \in \{1, 2, 3\}$). Next, we calculate the “triangular” variance of the realized average return at time $t+1$ as

$$TV_{t+1} = \sum_{l=1}^L r_{T,t+\frac{l}{L}}^2 + 2 \sum_{l=2}^L r_{T,t+\frac{l}{L}} r_{T,t+\frac{l-1}{L}}, \quad (6)$$

where L is the number of periods used in a sliding window. We also define realized variance of the returns to exchange rate j at time $t+1$ as

$$RV_{j,t+1} = \sum_{l=1}^L r_{j,t+\frac{l}{L}}^2 + 2 \sum_{l=2}^L r_{j,t+\frac{l}{L}} r_{j,t+\frac{l-1}{L}}, \quad j \in \{1, 2, 3\}. \quad (7)$$

The average “triangular” variance (ATV) and correlation (ATC) can be further written as

$$ATV_{t+1} = \frac{1}{3} \sum_{j=1}^3 RV_{j,t+1}, \quad (8)$$

$$ATC_{t+1} = \frac{1}{6} \sum_{j=1}^3 \sum_{i \neq j}^3 TC_{ij,t+1}, \quad (9)$$

where

$$TC_{t+1} = \frac{RV_{ij,t+1}}{\sqrt{RV_{i,t+1}} \sqrt{RV_{j,t+1}}}, \quad (10)$$

$$RV_{ij,t+1} = \sum_{l=1}^L r_{i,t+\frac{l}{L}} r_{j,t+\frac{l}{L}} + 2 \sum_{l=2}^L r_{i,t+\frac{l}{L}} r_{j,t+\frac{l-1}{L}}, \quad i, j \in \{1, 2, 3\}. \quad (11)$$

Note that ATV and ATC measures are the components of the “triangular” variance (TV), and this decomposition can be expressed as follows

$$TV_{t+1} = ATV_{t+1} \times ATC_{t+1}. \quad (12)$$

Two regressions will be of interest:

$$r_{TR,t+1} = \alpha + \beta TV_t + \epsilon_t, \quad (13)$$

$$r_{TR,t+1} = \alpha + \beta_1 ATV_t + \beta_2 ATC_t + \epsilon_t, \quad (14)$$

where $r_{TR,t+1}$ (triangular arbitrage returns) stands for the left-hand sides of Equation 3 and Equation 4, i.e., both routes for triangular arbitrage will be explored. Hence, $r_{TR,t+1}$ can be zero (triangular parity condition holds), and can take positive (arbitrage strategy profit) and negative (arbitrage strategy loss) values. Equation 14 separates the effects of ATV and ATC and we will establish the contribution of each component of TV in explaining and predicting triangular arbitrage returns. Our predictive regressions will employ lagged TV , ATV and ATC variables.

3.3. Limit Order Book Measures: Bid-Ask Spread and Center of Gravity

Having introduced the realized risk measures, we now turn to the role of limit order book in triangular arbitrage dynamics. The literature that describes the informativeness of the limit order book in equity markets is abundant, while such research in FX markets has been less intense, mainly due to the dispersed nature of the FX market and the unavailability of detailed high-frequency FX trading information.⁷ We are particularly interested in exploiting the shape of the limit order book which can be then used for predicting future prices. For example, Harris and Panchapagesan (2005) find that the limit order book is informative in revealing pending price changes. In a related paper, Cao *et al.* (2009) confirm the findings by Harris and Panchapagesan (2005) based on data gathered from the Australian Stock Exchange. Specifically, they document that the limit order book is informative in determining the value of an asset, as its contribution beyond the best bid and offer is 22%. Overall, the research findings reveal that lagged order book information is significantly correlated to future returns. In the same vein, Bloomfield *et al.* (2005) and Kaniel and Liu (2006) show that informed traders are more likely to favor limit orders over market orders. Kozhan and Salmon (2012) demonstrate the superiority of limit order book information in high-frequency out-of-sample FX rate forecasting and devising a profitable trading strategy. As a practical extension, we now utilize the limit order book to explain and predict triangular arbitrage profits.

The dynamic of the limit order book is such that at any point of time it contains a large number of orders over the bid and offer ranges. In general, these unrealized orders (i.e., price-quantity combinations) represent the aggregate FX market demand and supply schedules. To capture the structure of the limit order book in terms of both price and quantity, we use two measures – the

⁷Kozhan and Salmon (2012) provide an extensive literature review on the topic.

quantity-weighted average bid-ask spread and the novel measure we refer to as the *center of gravity quantity-weighted average bid-ask spread* – as well as the measure based on standard inside spread or the difference between the best bid and the best ask price.⁸ Our principal hypothesis is that the shape of the limit order book, when averaged across the three exchange rates, will have certain predictive power for future triangular arbitrage returns.

As before, if we substitute the three exchange rates (EUR/USD, EUR/JPY and USD/JPY) with ordinal numbers ($i = 1, 2, 3$), the average inside spread at time $t + 1$ can be written as

$$\overline{ispread}_{t+1} = \sum_{i=1}^3 \frac{S_{i,t+1}^{a,1} - S_{i,t+1}^{b,1}}{3}, \quad (15)$$

where $S_{i,t+1}^{a,1}$ and $S_{i,t+1}^{b,1}$ are the best ask and bid quotes of the i^{th} exchange rate, respectively. Therefore, superscript $(a, 1)$ stands for the ask price where price rank 1 represents that this is the best ask price (rank 2 is the 2nd best, etc.). In the same manner, superscript $(b, 1)$ stands for the best bid price with price rank 1. Figure 1 presents the location of the inside spread for the EUR/USD limit order book. The intuition behind this and other limit order book measures that we would like to capture shocks taking place in at least one of the markets. The shocks may cause price movements in the limit order book that violate the parity condition, which will be reflected in the change of the (average) measure. We conjecture that the larger the average inside spread, i.e., the distance between the bid and offer prices, the more difficult it becomes to profit from triangular arbitrage. Another motivation for using the spread as a predictor is that we observed an increased number of triangular parity violations during the times when the FX market is more active (i.e., when it is more liquid) in the three time zones.

Given the inside spread, it is important to emphasize that there could be a potential mismatch between the *real trading time* that the orders are executed in the EBS platform (with increments e.g., $\tau < 100\text{ms}$) and our constructed *sampling time* for which the incremental change between two consecutive observations is set to 100ms. While a database relying solely on real trading time observations is preferable, we notice that observations at frequencies higher than 100ms are heavily contaminated by certain data frictions, such as zero observations, extreme rapid drifts, constant transactions and noise (either market micro structure or distortion linked to speed of executions). As a consequence, any (potential) triangular arbitrage opportunity that is likely to be observed/detected at frequencies higher than 100ms may be severely exposed to such aforementioned frictions in the real time trading data. In order to avoid this problem and measure

⁸By using the information share measure from Hasbrouck (1995), Cao *et al.* (2009) show that the mid-quote and the quantity-weighted average mid-quote contribute to the price discovery by about 77%. Consequently, we consider these measures as our primary choice for the limit order book predictors of triangular arbitrage returns.

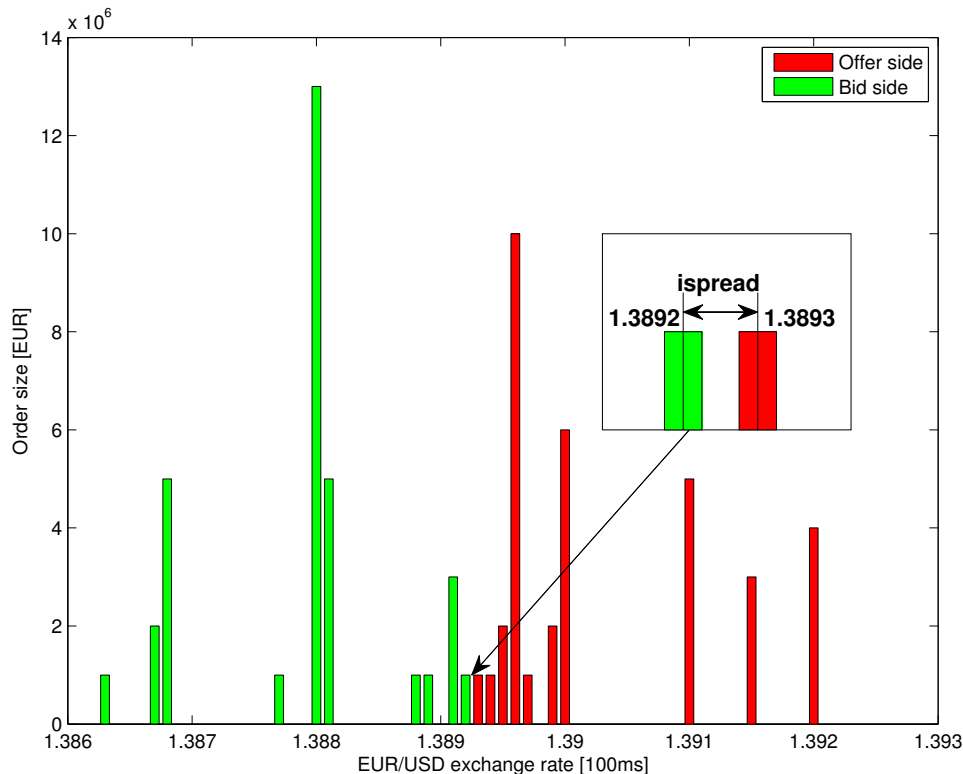


Figure 1: EXAMPLE: INSIDE SPREAD.

The data are taken by a random draw from the limit order book for the EUR/USD transactions at a given point in time at the highest frequency [100ms]. All ten levels on the bid and ask sides are visible with the height of the individual columns corresponding to the limit order size in the EUR currency.

triangular arbitrage with reasonable (yet still ultra high) frequency, we utilize instead the sampling time database which barely measures opportunities at frequencies $> 100ms$, but provides accuracy by limiting false detections.

Next, we use the complete limit order book information and define the quantity-weighted bid quote of an exchange rate i at time $t + 1$ as

$$qw_{i,t+1}^b = \frac{\sum_{j=1}^{10} S_{i,t+1}^{b,j} \times Q_{i,t+1}^{b,j}}{\sum_{j=1}^{10} Q_{i,t+1}^{b,j}}, \quad (16)$$

where j is the price rank or the level of the orders on the bid side and $Q_{i,t+1}^{b,j}$ is the corresponding

order size for the price $S_{i,t+1}^{b,j}$. On the ask side, we define the quantity-weighted ask quote of an exchange rate i at time $t + 1$ as

$$qwa_{t+1}^i = \frac{\sum_{j=1}^{10} S_{i,t+1}^{a,j} \times Q_{i,t+1}^{a,j}}{\sum_{j=1}^{10} Q_{i,t+1}^{a,j}}, \quad (17)$$

where the notation follows Equation 16. Based on Equations 16 and 17, we define the quantity-weighted average bid-ask spread as

$$\overline{qws\text{pread}}_{t+1} = \sum_{i=1}^3 \frac{qwa_{t+1}^i - qwb_{t+1}^i}{3}, \quad (18)$$

and also the quantity-weighted average mid-quote as

$$\overline{qwmidq}_{t+1} = \frac{1}{3} \sum_{i=1}^3 \frac{(qwa_{t+1}^i + qwb_{t+1}^i)}{2}. \quad (19)$$

Figure 2 shows the quantity-weighted ask quote, the quantity-weighted bid quote and the quantity-weighted mid-quote for the EUR/USD limit order book. These measures provide more information about the current limit order “pressure” on the price while accounting for the order size at each price level. In other words, they summarize all information contained in the order book that is relevant for future price movements. We average these measures across the three exchange rates from the triangular parity relationship.

The last measure we propose is a novel limit order book indicator that is inspired by the “center of gravity” concept from fuzzy logic. This predictor captures the most likely location of the bid and ask quotes by making the structure of the limit order book more “continuous” relative to the simple quantity-weighted approach. We define the center of gravity quantity-weighted bid quote of an exchange rate i at time $t + 1$ as

$$cogqwb_{t+1}^i = \frac{\int_{z^b} Q_{i,t+1}^b(z) z dz}{\int_{z^b} Q_{i,t+1}^b(z) dz}, \quad (20)$$

where z^b is the total area above the bid price levels defined by the shape of the bid side and $Q_{i,t+1}^b$ terms are the corresponding order sizes for price terms $S_{i,t+1}^b$. Then, we define the center of gravity

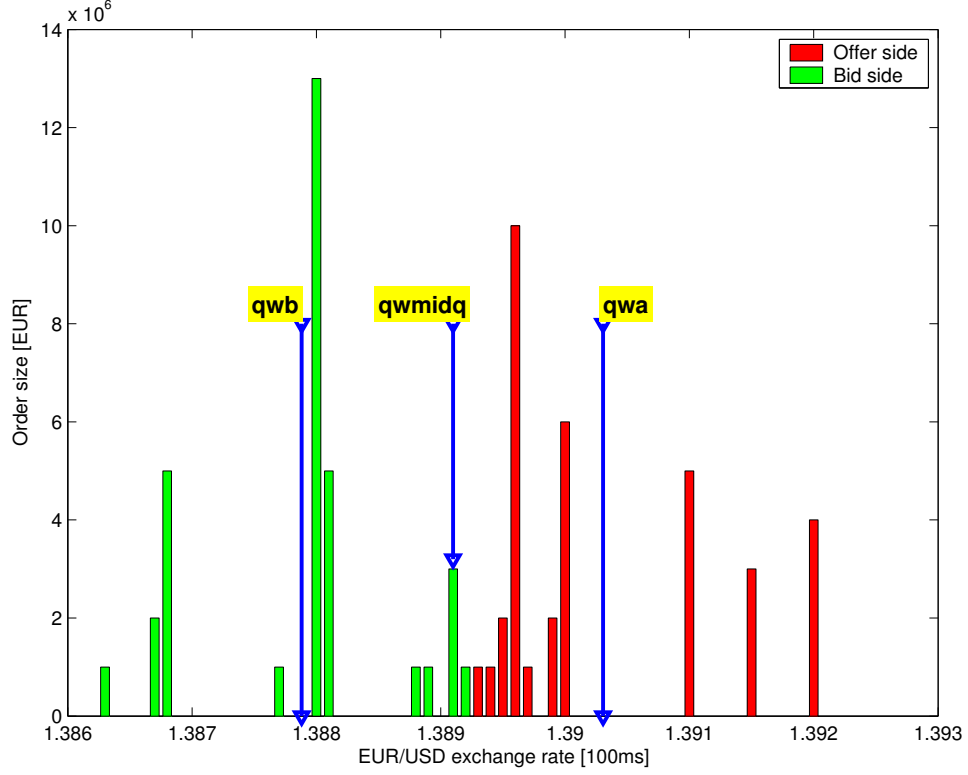


Figure 2: EXAMPLE: QUANTITY-WEIGHTED MEASURES.

The data are taken by a random draw from the limit order book for the EUR/USD transactions at a given point in time at the highest frequency [100ms]. All ten levels on the bid and ask sides are visible with the height of the individual columns corresponding to the limit order size in the EUR currency. The quantity-weighted ask quote (qwa), the quantity-weighted bid quote (qwb) and the quantity-weighted mid-quote ($qwmidq$) values for the snapshot of the EUR/USD limit order book are marked with arrows.

quantity-weighted ask quote of an exchange rate i at time $t + 1$ as

$$cogqwa_{t+1}^i = \frac{\int_{z^a} Q_{i,t+1}^a(z) z dz}{\int_{z^a} Q_{i,t+1}^a(z) dz}, \quad (21)$$

where z^a is the total area above the ask price levels defined by the shape of the ask side and $Q_{i,t+1}^a$ terms are the corresponding order sizes for price terms $S_{i,t+1}^a$. We can write the center of gravity

quantity-weighted average bid-ask spread as

$$\overline{cogqws\text{spread}}_{t+1} = \sum_{i=1}^3 \frac{cogqwa_{t+1}^i - cogqwb_{t+1}^i}{3}, \quad (22)$$

and the center of gravity quantity-weighted average mid-quote as

$$\overline{cogqwmidq}_{t+1} = \frac{1}{3} \sum_{i=1}^3 \frac{(cogqwa_{t+1}^i + cogqwb_{t+1}^i)}{2}. \quad (23)$$

The center of gravity concept is illustrated by Figure 3 which shows the center of gravity quantity-weighted ask quote, the center of gravity quantity-weighted bid quote and the center of gravity quantity-weighted mid-quote for the EUR/USD limit order book. Essentially, the center of gravity measure takes into account the whole area above the bid/ask price levels defined by the shape of the bid/ask sides (i.e., the integral of the area is taken). Intuitively, it should be a more accurate measure of the current limit order “pressure” on the price while accounting for the order size at each potential price level. As we will see later, the new measure, when averaged across the three exchange rates, is as successful as the simple quantity-weighted measure in capturing future triangular arbitrage returns.

4. Data

We use the latest generation of Electronic Broking Services (EBS) data called “Data Mine Level 5.0” from which we extract tick-by-tick FX transaction prices for the EUR/USD, EUR/JPY and USD/JPY exchange rates. EBS operates as an electronic limit order book and is used for global interdealer spot trading. It is dominant and most representative for the EUR-USD and USD-JPY currency trading, whereas the GBP-USD currency pair is traded primarily on Reuters. The data are recorded for ten best bid and ten best offer prices for each exchange rate over 24 hours, based on GMT time. The best bid is the highest bid price in the EBS market, while the best offer is the lowest offer price in the EBS market at the time, regardless of credit. EBS provides ten layers of prevalent (“transactable”) best bid and ask quotes as well as the corresponding order sizes. The direction of each trade is known and transaction costs are directly measured by the bid-ask spread. To demonstrate the robustness of our analysis, we choose the following random non-overlapping time periods with the observation frequency of 100 milliseconds (1/10th of a second): November 1-14, 2010, February 21-27, 2011, April 4-10, 2011, October 3-16, 2011, excluding weekends. The weeks are sampled so that the findings would not be affected by seasonal market fluctuations (e.g.,

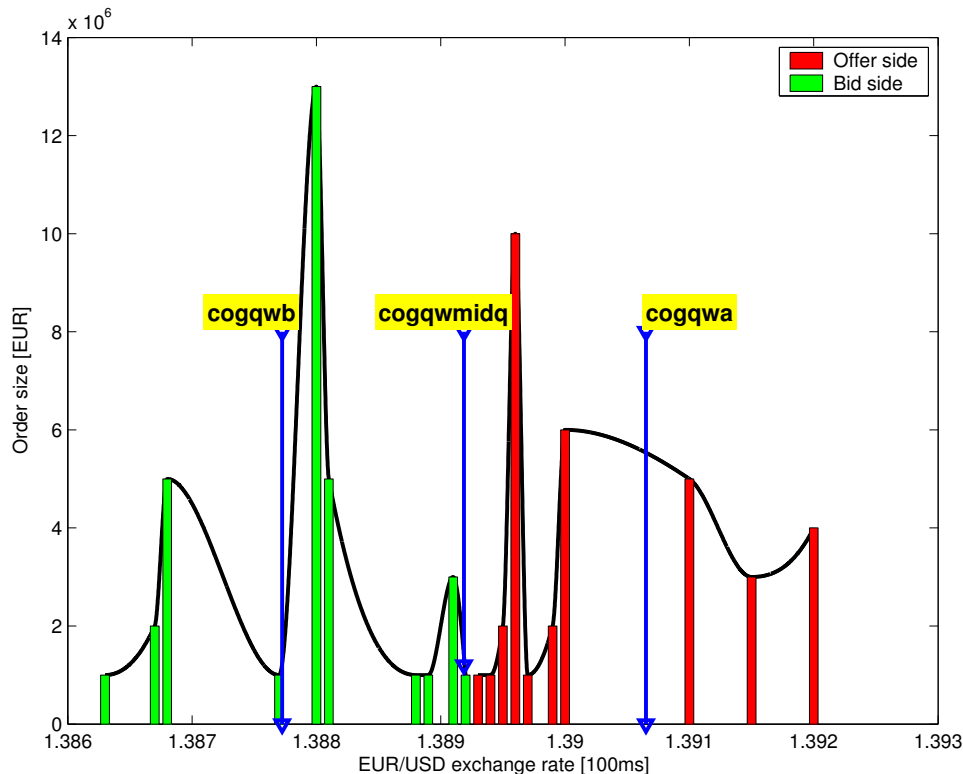


Figure 3: EXAMPLE: CENTER OF GRAVITY QUANTITY-WEIGHTED MEASURES.

The data are taken by a random draw from the limit order book for the EUR/USD transactions at a given point in time at the highest frequency [100ms]. All ten levels on the bid and ask sides are visible with the height of the individual columns corresponding to the limit order size in the EUR currency. The center of gravity quantity-weighted ask quote (*cogqwa*), the center of gravity quantity-weighted bid quote (*cogqwb*) and the center of gravity quantity-weighted mid-quote (*cogqwmidq*) values for the snapshot of the EUR/USD limit order book are marked with arrows.

Christmas/holiday season). Also, in order to minimize the impact of the Euro crisis (in February, 2010, hedge funds took the largest bearish bet against the Euro in the history), the sample starts in late 2010, and then covers mostly 2011. Each day contains about 25 million lines of data (quotes and transactions) for all exchange rates. Orders in the EBS market are submitted in units of millions of the base currency.⁹ For instance, if we consider EUR/USD prices, the quoted price is

⁹This means that the minimum order size is 1,000,000 EUR, USD or JPY, which may create an integer problem and impact traders to have some residual position exposure. The residual will always be smaller than a million units and can be cleared when it reaches one million units. Hence, the minimum order size feature of the EBS market may reduce triangular arbitrage returns. However, traders could deal with this problem by using multiple trading channels

the amount of local currency (USD) that is required to purchase one unit of the base currency (EUR).

In contrast to equity limit order books, EBS platform has a unique configuration that constrains buyers and sellers to form a bilateral relationship based on credit exposure. Consequently, in cases when there is no specific bilateral credit between parties, the trade is unlikely to be executed which may in turn generate negative (or zero) spread. As a part of our data adjustment process, we removed the occurrences of all negative and zero spreads prior to the construction of our realized proxies and carrying out the estimations. In testing for triangular arbitrage, this refinement in turn allows us to acquire robustness with respect to presence of negative/zero spread or similar trade limits associated with platform-specific rules.¹⁰

By testing for Equations 3 and 4, we find on average about 100 daily triangular arbitrage opportunities over the November 1-14, 2010 time period. Both parity equations contribute roughly equally to the violations of the parity condition. The average daily returns from triangular arbitrage are: 0.5 bps (Equation 3) and 0.75 bps (Equation 4). For the second two-week time period (October 3-16, 2011), the average daily number of arbitrage opportunities decreases to about 80. The contribution of both parity equations is again roughly equal. The average daily return from Equation 3 for this period is 0.56 bps, while it is 0.62 bps based on Equation 4. The weighted average return thus decreases in the second period. The arbitrage parity violations are very short-lived and last between 100-500 milliseconds.

One important feature of triangular arbitrage opportunities is that the shifts in exchange rates open a profit that may be as large as 1 bps, but also quite small (below 0.5 bps). This opportunity may be immediately eroded to zero in the next 100-millisecond time interval or last for as long as 500 milliseconds while it gets gradually arbitrated away. As we only sample at the 100-millisecond frequency, we can observe arbitrage profits for each successive time period, which can be single (100 milliseconds) or multiple periods (more than 100 milliseconds). When we calculate the average figures in Table 1, we average across various instances of profit durations. As a note of caution, it is important to mention that, due to the 100-millisecond frequency limitation, certain arbitrage

that process trades below 1 million units and will normally provide competitive quotes (Bjønnes and Longarela, 2014). Another possibility for FX traders is to maintain large inventories of currencies and/or to accumulate larger positions over time in an attempt to use a favorable future FX rate and clear them. Generally speaking, this is a known problem in triangular arbitrage and it is beyond the scope of the paper to investigate the practical trading strategies used for clearing the residual positions.

¹⁰It is of course worth emphasizing that in cases when involved traders cannot trade in certain situations, the associated impact on the assessment of triangular arbitrage opportunities is expected to remain fairly marginal, see e.g., the recent evidence provided by Gould *et al.* (2017) for the Hotspot FX platform. Focusing on this particular dimension, future research may investigate the role of wavelet analysis in characterizing arbitrage opportunities conditional on the given *asymmetric* credit structures (i.e., with respect to zero/positive/negative spreads).

Data range	N	Mean	Stdev	Duration
November 1-14, 2010	96	0.63	0.20	170
February 21-27, 2011	34	0.34	0.21	110
April 4-10, 2011	120	0.46	0.09	280
October 3-16, 2011	82	0.59	0.38	130

Table 1: SUMMARY STATISTICS OF TRIANGULAR ARBITRAGE PROFITS.

Notes: For each time period (“Data range”), the table presents the daily averages of the following statistics: the number of triangular arbitrage opportunities (“N”), triangular arbitrage strategy profits expressed in bps (“Mean”), standard deviation of triangular arbitrage strategy profits expressed in bps (“Stdev”) and duration of triangular arbitrage opportunities expressed in milliseconds (“Duration”).

opportunities may in fact be shorter and only survive for a few nanoseconds.¹¹

Table 1 presents the summary statistics for triangular arbitrage profits for each of the four time periods that we consider in our analysis. Interestingly, we observe a large increase in the average daily number of triangular arbitrage opportunities as well as a decrease in the standard deviation of returns over the April 4-10, 2011 period. It appears that this particular week involved more frequent and relatively stable profits to triangular arbitrage. Also, the average duration of triangular arbitrage opportunities increased during this period, but the same is not observed for the magnitude of the profits (0.46 bps). November 1-14, 2010 and October 3-16, 2011 represent relatively normal market regimes when we observe reduction in the average number of arbitrage opportunities and their profitability over time, from 2010 to 2011. February 21-27, 2011 and April 4-10, 2011 represent abnormal market regimes (before and after the EBS tick change). Regime changes are directly responsible for the fluctuations in mean values. In Section 6.3, we will study these results in more detail.¹²

5. Estimation and Results: Wavelet-based Regressions

After discussing the triangular arbitrage strategy and our predictors, we now introduce our estimation scheme that relies on wavelet-approach in regression analysis. We proceed as follows. Following Gençay and Gradojevic (2011), we first present the statistical problem behind the ordinary least squares estimation. We then present our correction which relies on wavelet-based

¹¹Transaction costs in the FX market are accounted for by the bid-ask spreads that are present in an interbank trading platform such as EBS. In regards to potential commissions paid to FX brokers, we assume that banks trade directly on EBS and do not have to use any brokers. In general, the commission FX dealers/brokers might apply would diminish the profits to triangular arbitrage.

¹²Caporin *et al.* (2015) and Mahmoodzadeh and Gençay (2016) provide a more detailed treatment of the limit order book data from EBS.

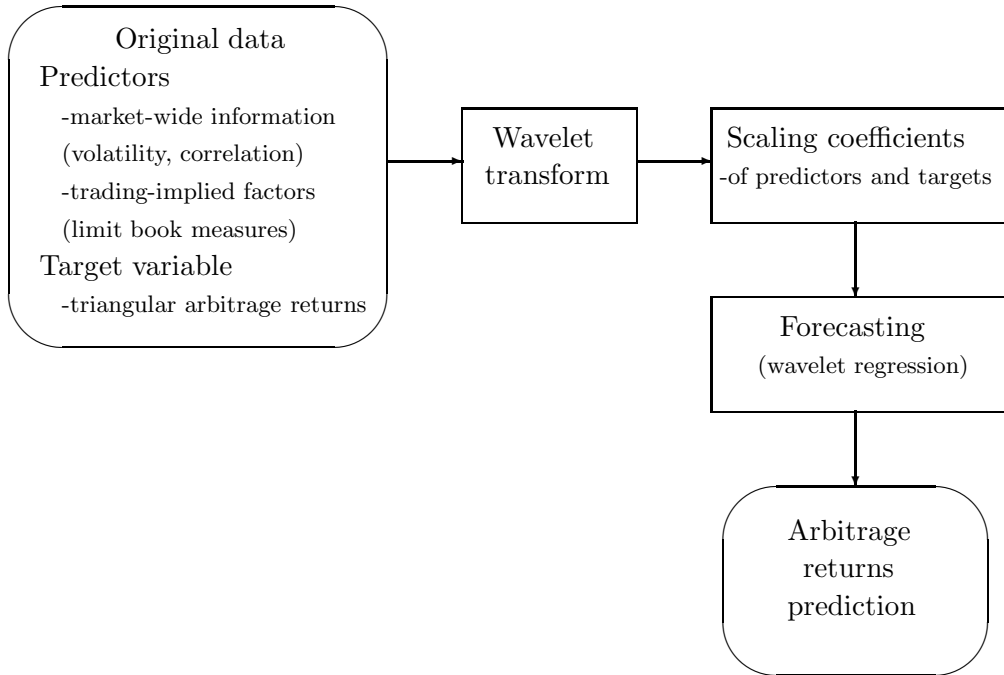


Figure 4: FLOWCHART OF RESEARCH FRAMEWORK.

Notes: Wavelet transform filters out the measurement noise and uses the filtered prediction target (scaling coefficients of triangular arbitrage returns) and predictors (scaling coefficients of ATV , ATC , TV and order book measures) in OLS to obtain unbiased and consistent estimators at ultra-high-frequency. The resulting predictive linear regressions are one-step-ahead forecasting exercises.

regression approach that directly extends the work of Gençay and Gradojevic (2011) to empirical ultra-high-frequency analysis. We illustrate the advantage of using our proposed wavelet-based estimation with real data.¹³ It should also be noted that our wavelet-based regression methodology does not formally consider causal relationships and that such explorations are outside the paper's scope. We provide the technical details with respect to estimation in the Appendix.

¹³For brevity, we do not report our Monte Carlo results with respect to the finite sample properties of our estimators. These results are available upon request.

5.1. Persistence of Predictors: The Robust Estimation with Wavelets

As in Gençay and Gradojevic (2011), we consider a linear regression model

$$y_t^* = \alpha + \beta x_t^* + \epsilon_t^*, \quad \epsilon_t^* \sim iid(0, \sigma_{\epsilon^*}^2) \quad (24)$$

where y_t^* and x_t^* are unobserved persistent processes, ϵ_t is identically and independently distributed (*iid*) with variance $\sigma_{\epsilon^*}^2$. The observables are

$$x_t = x_t^* + v_{1t} \quad \text{and} \quad y_t = y_t^* + v_{2t} \quad (25)$$

where $v_{1t} \sim iid(0, \sigma_{v1}^2)$ and $v_{2t} \sim iid(0, \sigma_{v2}^2)$ are measurement errors.

In the presence of persistent predictors subject to measurement noise, the OLS estimator is downward biased and asymptotically inconsistent. A conventional way to deal with the inconsistency of the OLS estimator is to use instrumental variables (IV) estimation. To tackle the estimation problem, we propose filtering out the measurement noise and use the filtered regressand and regressors in OLS to obtain unbiased and consistent estimators at ultra-high-frequency. Extending Gençay and Gradojevic (2011), we apply the wavelet method to both y_t and x_t measured at ultra-high-frequency, and regress the scaling coefficients of S_{y_t} onto S_{x_t} . We run the following *wavelet-based arbitrage regression*

$$S_{y_t} = \alpha + \beta S_{x_t} + \tilde{\epsilon}_t, \quad \tilde{\epsilon}_t \sim iid(0, \sigma_{\tilde{\epsilon}}^2) \quad (26)$$

instead of Equation 24. Since noise terms are left behind in the wavelet coefficients, the scaling coefficients will provide unbiased and consistent coefficient estimators without any instruments added to the regression. Gençay and Gradojevic (2011) show that the wavelet approach is particularly effective when the regressand and the regressor are persistent processes. Without the correction, we find all the variables in our predictive regressions stationary, but highly persistent and, due to market microstructure noise, possibly subject to measurement errors. Figure 4 depicts the flow diagram of our research design with real data.

5.2. Persistence of Predictors: Illustration and patterns with real data

The above approach filters out the measurement noise and helps with tackling persistent variables. To illustrate this, we plot the autocorrelation function (ACF) for the *ATV* indicator and for its corresponding Haar wavelet scaling coefficients on November 1, 2010 (Figure 5). The bottom panel of Figure 5 shows that the ACF for the Haar wavelet scaling coefficients exhibits substantially less

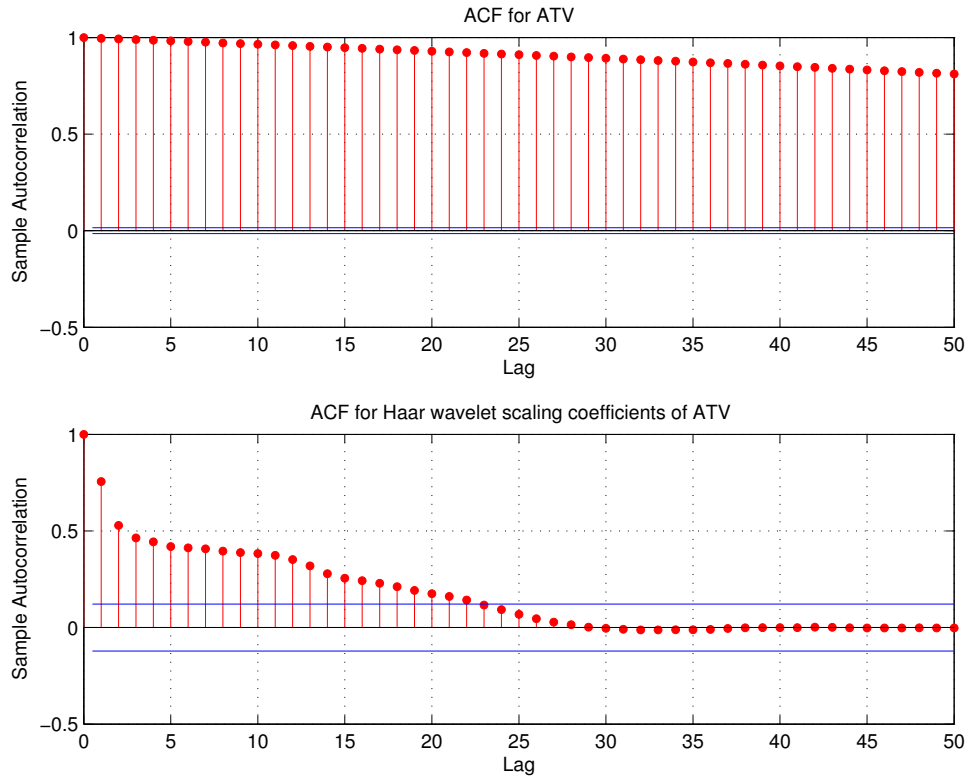


Figure 5: SAMPLE AUTOCORRELATION FUNCTIONS FOR *ATV*.

The data are taken at the highest available frequency on November 1, 2010. Top panel: autocorrelation function (ACF) for the *ATV* indicator (N=17310). Bottom panel: autocorrelation function (ACF) for the Haar wavelet scaling coefficients (maximum overlap discrete wavelet transform, level of decomposition is 6) estimated for the *ATV* indicator (N=271).

persistence than we find in the top panel of Figure 5. In addition, the wavelet transform of all our regressands and the regressors improves the explanatory power of the predictive regressions. We run the regression from Equation 26, i.e., scaling coefficients for the triangular arbitrage returns are regressed on scaling coefficients for the predictor variables. The results are shown in Tables 2 and 3. Clearly, for the period Nov. 1-14, 2010, on average, the estimates suggest that all the predictive variables are useful in forecasting triangular arbitrage returns. The \overline{R}^2 figures are substantially larger and the coefficients do not show any significant fluctuations.¹⁴

In the above illustration, we removed six wavelet scales in order to be consistent with Gençay and Gradojevic (2011) who documented by using simulation evidence that higher levels of decomposition

¹⁴The average results for the October 3-16, 2011 period are similar and, for brevity, we do not present the tables.

Nov. 1-14, 2010	Eq 13(2)	Eq 13(2)	Eq 14(3)	Eq 14(3)
Constant	1.999 (0.000)	1.999 (0.000)	1.999 (0.000)	1.999 (0.000)
TV_t	-36.82 (0.000)		-21.18 (0.000)	
ATC_t		-0.001 (0.000)		-0.001 (0.000)
ATV_t		-18.37 (0.000)		-10.04 (0.000)
\overline{R}^2	[0.129]	[0.199]	[0.094]	[0.142]

Table 2: PREDICTIVE POWER OF SCALING COEFFICIENTS FOR ATV , ATC AND TV (NOVEMBER 1-14, 2010).

Notes: The table presents the daily averages for ordinary least squares regression results for one-step-ahead forecasting of scaling coefficients for triangular arbitrage returns (Equations 13 and 14): “Eq 13(2)” with “2” in the parentheses denotes that Equation 3 was used for the calculation of triangular returns (i.e., $r_{TR,t+1}$ from the first triangular arbitrage route) and “Eq 14(3)” with “3” in the parentheses denotes that Equation 4 was used for the calculation of triangular returns (i.e., $r_{TR,t+1}$ from the second triangular arbitrage route). The numbers in parentheses are Newey and West (1987) p-values with ten lags for the estimates and the numbers in square brackets are the \overline{R}^2 values for each predictive regression. The regressors are scaling coefficients for the triangular variance (TV), average triangular variance (ATV) and correlation (ATC), defined in Equation 8 and Equation 9, respectively.

increase the precision of the MODWT (Haar) estimates of the slope coefficient ($\hat{\beta}$). But, at the same time, it is important to be careful with the terminology and stress that noise filtering refers to only taking out one or at most two scales of wavelet coefficients. In Figures 6 and 7, we show that the extraction of memory is similar for one and three scales, while most of the persistence is taken out in scale one. Both figures resemble Figure 5.¹⁵

¹⁵It is thus important to note that market microstructure noise is very likely to distort the measurement of the proxies of our realized quantities. Before conducting our estimations, we aim at handling the potential frictions altered by noise in two ways. First, we pre-filtered our raw data thoroughly by removing excessive amount of zero observations, consecutive constant price levels and diurnal effects. While we do not use robust-noise measures (leaving this as future research), we clean our database from potential frictions which in turn limit the effect of noise observed at ultra high frequencies. Secondly, from a technical perspective, we use wavelet methodology to obtain estimates that are reliable even in the presence of microstructure noise in the ultra high frequency data. Our wavelet (noise) filtering approach serves this purpose. The advantage of the wavelet method in shrinking the effect of noise is in fact related to the *denoising* mechanics of wavelets, which particularly holds for the Haar wavelet (compared to smooth wavelet), consistent with the asymptotic results documented by Fan and Wang (2007). The estimates without wavelet transform are very likely to be contaminated with the noise whereas such effects are asymptotically negligible under wavelets with certain partition numbers. Consequently, our estimates are relatively less subject to the effects of microstructure noise. Along these lines, future research can develop the formal statistical inference and estimation scheme for our predictors in the presence of noise and other forms of frictions.

Nov. 1-14, 2010	$\overline{qwspread}_t$	$\overline{cogqwspread}_t$	$r_{TRqw,t}^i$	$r_{TRcoq,t}^i$	$[\overline{R}^2]$
$r_{TR,t+1}^1$ ($i = 1$)					
β_{1i}, β_{2i}		-0.000		0.042	[0.286]
(p-value)		(0.000)		(0.000)	
β_{1i}, β_{2i}	-0.001		0.110		[0.398]
(p-value)	(0.000)		(0.000)		
$r_{TR,t+1}^2$ ($i = 2$)					
β_{1i}, β_{2i}		-0.000		0.029	[0.246]
(p-value)		(0.000)		(0.000)	
β_{1i}, β_{2i}	-0.001		0.052		[0.339]
(p-value)	(0.000)		(0.000)		

Table 3: PREDICTIVE POWER OF SCALING COEFFICIENTS FOR LIMIT ORDER BOOK MEASURES (NOVEMBER 1-14, 2010; DAILY AVERAGES).

Notes: $r_{TR,t+1}^i$ (scaling coefficients for triangular arbitrage returns) stands for the left-hand sides of Equation 3 ($i=1$) and Equation 4 ($i=2$). The predictors are the scaling coefficients for quantity-weighted average bid-ask spread ($\overline{qwspread}$), the scaling coefficients for center of gravity quantity-weighted average bid-ask spread ($\overline{cogqwspread}$), scaling coefficients for triangular arbitrage returns obtained from the quantity-weighted average bid and ask prices ($r_{TRqw,t}^i$), and scaling coefficients for triangular arbitrage returns obtained from the center of gravity quantity-weighted average bid and ask prices ($r_{TRcoq,t}^i$). The numbers in parentheses are Newey and West (1987) p-values (p-value) with ten lags for the estimates of β_i , and the numbers in square brackets are the \overline{R}^2 values of each predictive regression.

5.3. Persistence of Predictors: Illustration and patterns with simulated data

Typically, persistence refers to the low-frequency part of the spectrum. Moreover, we argue that the presence of jumps and structural breaks at medium to high frequencies may create an “illusion” of low-frequency persistence. In fact, when we remove such high-frequency phenomena, significant degree of persistence at low frequencies is reduced, although persistence in the medium- to high-frequency part of the spectrum tends to be preserved.

To illustrate that wavelets are able capture such medium- to high-frequency persistence reflected in jumps (or structural breaks), we simulate two processes (N=10,000): 1) ARMA(1,1) with jumps, where $\phi=0.95$ and $\theta=0.5$, and 2) fractionally differenced process - ARFIMA(1,d,1) with $\phi=0.95$, $\theta=0.5$, $d=0.8$. The first process is supposed to represent persistency at medium to high frequencies with jumps, while the second one captures short-run persistence and long-memory.

Figures 8 and 9 display our findings. It is fairly noticeable that ACFs look persistent without “wavelet filtering”, whereas, after six scales are removed, the decay in ACFs is exponential in ARMA(1,1) and hyperbolic in ARFIMA(1,d,1), which is similar to patterns revealed by Figure 5.

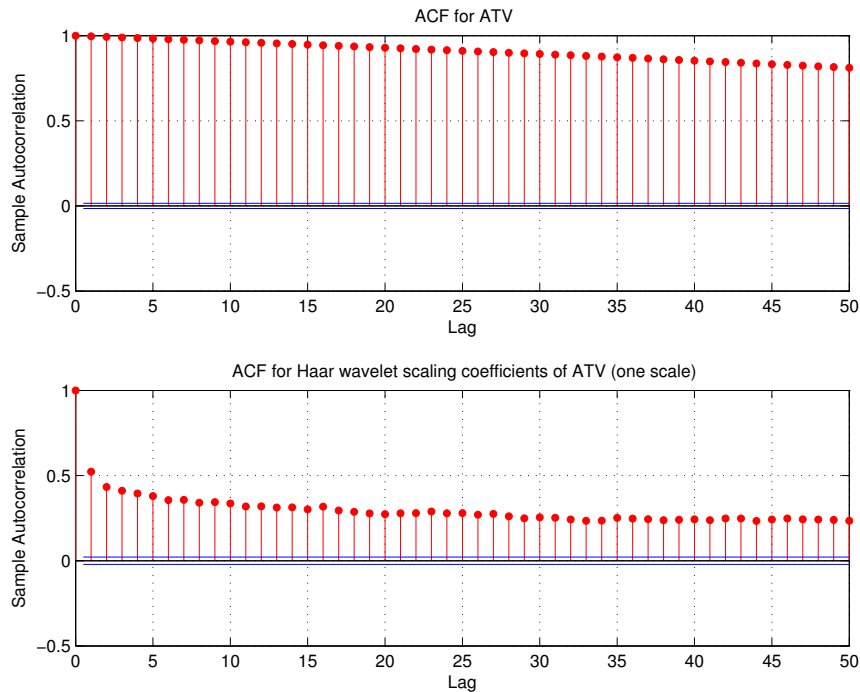


Figure 6: SAMPLE AUTOCORRELATION FUNCTIONS FOR *ATV* (LEVEL 1).

The data are taken at the highest available frequency on November 1, 2010. Top panel: autocorrelation function (ACF) for the *ATV* indicator (N=17310). Bottom panel: autocorrelation function (ACF) for the Haar wavelet scaling coefficients (maximum overlap discrete wavelet transform, level of decomposition is 1) estimated for the *ATV* indicator (N=8655).

6. Empirical Results

This section examines the correlation dynamics, robust prediction of triangular arbitrage and the effect of decimal pip pricing on the predictions.

6.1. Correlation Dynamics

To explore further the relationship between the proposed predictors and triangular arbitrage returns, Table 4 presents correlation coefficients for the *ATC*, *ATV* and *TV* variables. Significance probabilities under the null hypothesis of no correlation are $p=0.000$ for all cells in the correlation matrix, i.e., all predictors exhibit statistically significant correlation coefficients both contemporaneously and lagged. The contribution of the *ATC* measure in explaining and predicting triangular arbitrage returns is much smaller relative to the *ATV* predictor variable. Also, all the measures are

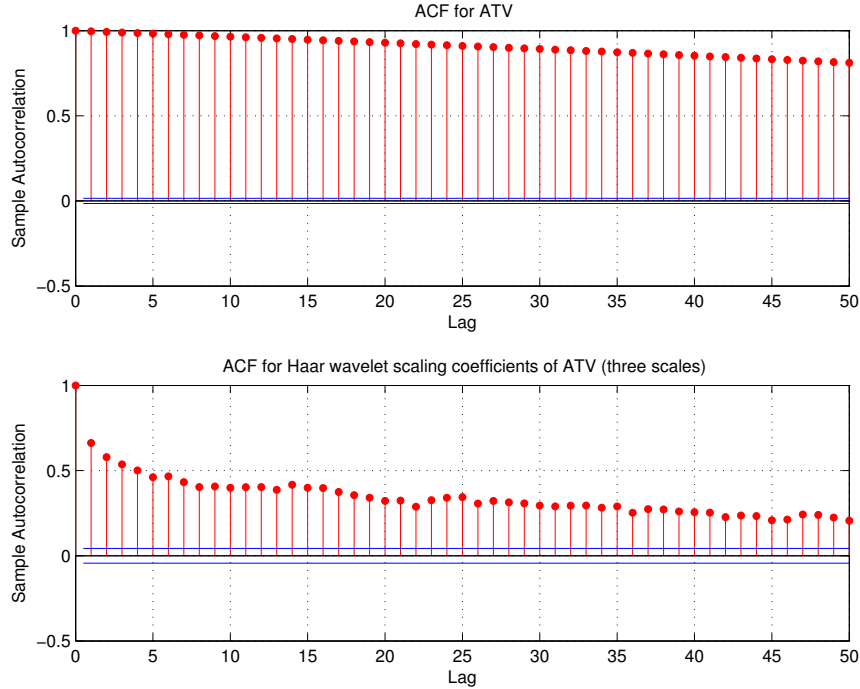


Figure 7: SAMPLE AUTOCORRELATION FUNCTIONS FOR ATV (LEVEL 3).

The data are taken at the highest available frequency on November 1, 2010. Top panel: autocorrelation function (ACF) for the ATV indicator ($N=17310$). Bottom panel: autocorrelation function (ACF) for the Haar wavelet scaling coefficients (maximum overlap discrete wavelet transform, level of decomposition is 3) estimated for the ATV indicator ($N=2164$).

negatively correlated to triangular arbitrage returns. This suggests that triangular arbitrage profits are more likely when the average price volatility and the average interaction among the exchange rates are low. We observe a relatively weak correlation between triangular arbitrage returns $r_{TR,t+1}^1$ and $r_{TR,t+1}^2$ (0.07), which points to differential nature of the two arbitrage strategies.

Next, we construct the correlation matrix for triangular arbitrage returns and the limit order book measures. The basic predictors that we use are the average inside spread (Equation 15) and the quantity-weighted average bid-ask spread (Equation 18). In addition, we calculate the benchmark measures based on substituting the bid and ask quotes in Equations 3 and 4 for the quantity-weighted bid quote (qwb_{t+1}^i ; $i = 1, 2, 3$) from Equation 16 as well as the quantity-weighted ask quote (qwa_{t+1}^i ; $i = 1, 2, 3$) from Equation 17. We denote these measures $r_{TRqw,t}^1$ and $r_{TRqw,t}^2$ for Equations 3 and 4, respectively. The intuition behind these predictors is that the quantity-weighted bid and ask quotes may represent future realizations of the actual, transactable bid and ask quotes

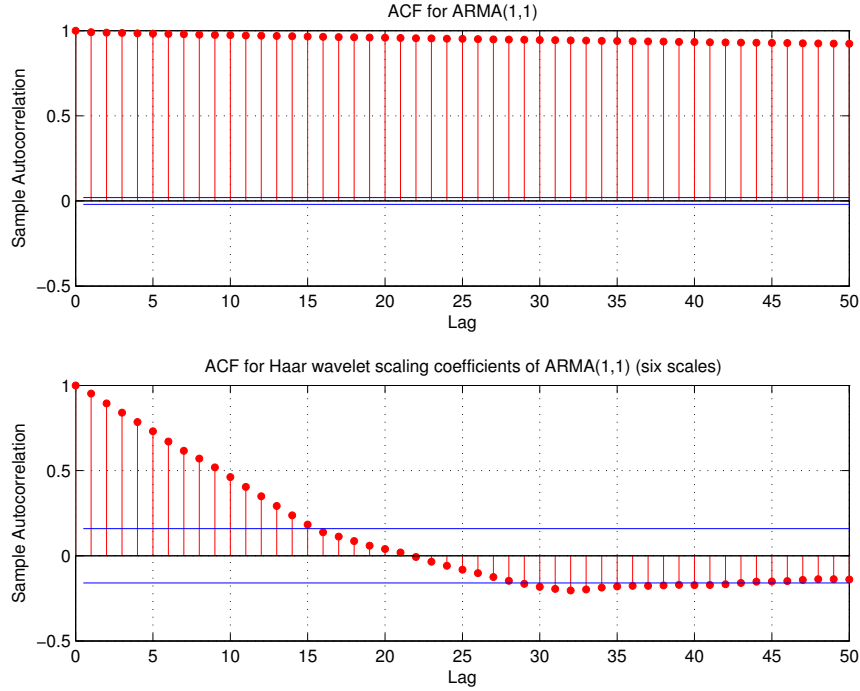


Figure 8: SAMPLE AUTOCORRELATION FUNCTIONS FOR ARMA(1,1).

The data are simulated from ARMA(1,1) with jumps (five jumps are located at points 2,000; 4,000; 5,000; 7,000; and 9,000; their magnitude is five standard deviations from the mean value), where $\phi=0.95$ and $\theta=0.5$. Top panel: autocorrelation function (ACF) for the simulated ARMA(1,1) process (N=10,000). Bottom panel: autocorrelation function (ACF) for the Haar wavelet scaling coefficients (maximum overlap discrete wavelet transform, level of decomposition is 6) estimated for the simulated ARMA(1,1) process (N=157).

in the order book. In turn, the triangular arbitrage returns received by using the quantity-weighted bid and ask quotes represent forecasts of future triangular arbitrage returns.

Table 5 reveals a strong negative contemporaneous correlation between the standard inside spread measure and triangular arbitrage returns (-0.57 and -0.52). Similarly, triangular arbitrage profits diminish as the quantity-weighted average bid-ask spread among the exchange rates widens. The corresponding correlation coefficients are -0.40 and -0.41. In a predictive setting, however, the lagged quantity-weighted average bid-ask spread becomes more dominant (correlation coefficients: -0.33 and -0.34) while the lagged average inside spread displays weaker correlation coefficients with $r_{TR,t+1}^1$ (-0.30) and $r_{TR,t+1}^2$ (-0.31). The two new measures appear to be the most useful in forecasting triangular arbitrage returns: the correlation coefficient between $r_{TRqw,t}^1$ and $r_{TR,t+1}^1$ is

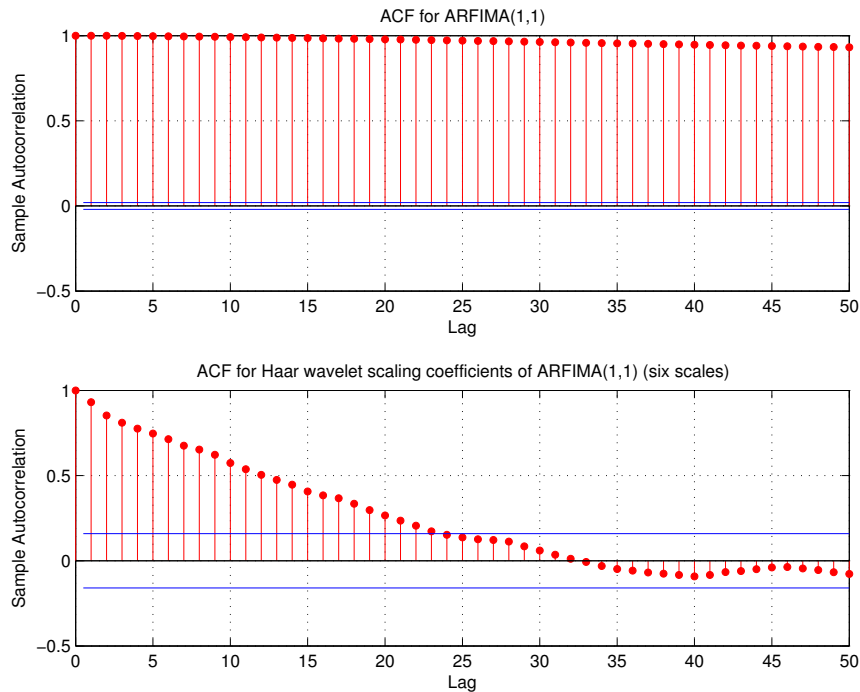


Figure 9: SAMPLE AUTOCORRELATION FUNCTIONS FOR ARFIMA(1,1).

The data are simulated from ARFIMA(1,d,1), where $\phi=0.95$, $\theta=0.5$, $d=0.8$. Top panel: autocorrelation function (ACF) for the simulated ARFIMA(1,d,1) process (N=10,000). Bottom panel: autocorrelation function (ACF) for the Haar wavelet scaling coefficients (maximum overlap discrete wavelet transform, level of decomposition is 6) estimated for the simulated ARFIMA(1,d,1) process (N=157).

0.36, and the corresponding figure for $r_{TRqw,t}^2$ and $r_{TR,t+1}^2$ is 0.35. Based on the above findings, our predictive regressions will not utilize the average inside spread and the quantity-weighted average bid-ask spread will be used instead.¹⁶

6.2. Robust Prediction of Triangular Arbitrage

To carry out robust prediction, we first test whether TV , ATV and ATC can provide insight into forecasting triangular arbitrage returns. We run linear regressions from Equations 13 and 14, and report our findings in Tables 6 and 7. The results indicate statistically significant forecast ability of the predictors. In summary, higher average triangular variance and average triangular correlation

¹⁶The correlations based on the center of gravity quantity-weighted average measures are similar to Table 5. For brevity reasons, we do not include another table to this section. It is available upon request.

Nov. 1-14, 2010	$r_{TR,t+1}^1$	$r_{TR,t+1}^2$	TV_{t+1}	ATV_{t+1}	ATC_{t+1}	TV_t	ATV_t	ATC_t
$r_{TR,t+1}^1$	1							
$r_{TR,t+1}^2$	0.07	1						
TV_{t+1}	-0.36	-0.23	1					
ATV_{t+1}	-0.35	-0.22	0.99	1				
ATC_{t+1}	-0.04	-0.03	0.11	0.04	1			
TV_t	-0.36	-0.23	0.99	0.99	0.11	1		
ATV_t	-0.35	-0.22	0.99	0.99	0.04	0.99	1	
ATC_t	-0.04	-0.03	0.11	0.04	0.99	0.11	0.04	1

Table 4: CORRELATIONS BETWEEN TRIANGULAR ARBITRAGE RETURNS AND PREDICTORS (NOVEMBER 1-14, 2010).

Notes: The table reports the average daily correlation coefficients between triangular arbitrage returns ($r_{TR,t+1}^i$; $i=1,2$ indices stand for the left-hand sides of Equation 3 and Equation 4, respectively) and the average triangular variance (ATV) and correlation (ATC), and triangular variance (TV) measures.

predict lower returns from triangular arbitrage. Put differently, high average exchange rate volatility makes triangular arbitrage profits more elusive. Similarly, triangular arbitrage opportunities require low average correlation across the three exchange rates from the triangular parity condition. These findings are intuitive and in accord with the evidence from Cenedese *et al.* (2014) that focus on gains from carry trade strategies. However, in contrast to Cenedese *et al.* (2014), we find that average triangular variance is a substantially better predictor of triangular profits than average triangular correlation. The latter appears to provide a smaller contribution to total triangular variance and thereby to the predictability of arbitrage profits. It is noticeable that the predictors are more successful in predicting triangular arbitrage returns in 2010 relative to 2011, in line with the adaptive market efficiency.¹⁷

In terms of the FX market microstructure lessons from the above findings, we conjecture that high average volatility creates a risky trading environment that reduces the likelihood of execution of potentially large trades (or, more generally, additional shocks) in any single currency from the triangular parity condition. Naturally, in less volatile markets, market shocks can have larger impacts and produce shifts in one of the exchange rate quotes, which may result in an arbitrage opportunity. As we have shown before, these opportunities are quickly exhausted by high-frequency traders that are able to react at the millisecond frequency. Furthermore, high average correlation among exchange rates can be interpreted as a relatively stable trading environment where the three exchange rates move in a synchronous fashion. This exchange rate quoting regime indicates the lack of potential for price shifts that open up arbitrage opportunities.

¹⁷Unfortunately, we did not have access to any other years of the EBS data at the ultra-high-frequency.

Nov. 1-14, 2010	$r_{TR,t+1}^1$	$r_{TR,t+1}^2$	$\overline{qwspread}_{t+1}$	$\overline{ispread}_{t+1}$	$r_{TRqw,t+1}^1$	$r_{TRqw,t+1}^2$	$\overline{qwspread}_t$	$\overline{ispread}_t$	$r_{TRqw,t}^1$	$r_{TRqw,t}^2$
$r_{TR,t+1}^1$	1									
$r_{TR,t+1}^2$	0.07	1								
$\overline{qwspread}_{t+1}$	-0.40	-0.41	1							
$\overline{ispread}_{t+1}$	-0.57	-0.52	0.53	1						
$r_{TRqw,t+1}^1$	0.46	0.30	-0.73	-0.49	1					
$r_{TRqw,t+1}^2$	0.39	0.46	-0.75	-0.49	0.48	1				
$\overline{qwspread}_t$	-0.33	-0.34	0.74	0.43	-0.67	-0.65	1			
$\overline{ispread}_t$	-0.30	-0.31	0.44	0.46	-0.41	-0.40	0.53	1		
$r_{TRqw,t}^1$	0.36	0.29	-0.66	-0.41	0.73	0.48	-0.73	-0.49	1	
$r_{TRqw,t}^2$	0.27	0.35	-0.65	-0.38	0.47	0.68	-0.75	-0.49	0.48	1

Table 5: CORRELATIONS BETWEEN TRIANGULAR ARBITRAGE RETURNS AND ORDER BOOK PREDICTORS (NOVEMBER 1-14, 2010).

Notes: The table reports the average daily correlation coefficients between triangular arbitrage returns ($r_{TR,t+1}^i$; $i=1,2$ indices stand for the left-hand sides of Equation 3 and Equation 4, respectively) and the following variables: quantity-weighted average bid-ask spread ($\overline{qwspread}$), inside spread ($\overline{ispread}$), left-hand side of Equation 3 when bid and ask prices are calculated by Equations 16 and 17 ($r_{TRqw,t+1}^1$) and left-hand side of Equation 4 when bid and ask prices are calculated by Equations 16 and 17 ($r_{TRqw,t+1}^2$).

Nov. 1-14, 2010	Eq 13(2)	Eq 13(2)	Eq 14(3)	Eq 14(3)
Constant	0.999 (0.000)	0.999 (0.000)	0.999 (0.000)	0.999 (0.000)
TV_t	-35.38 (0.000)		-20.62 (0.000)	
ATC_t		-0.011 (0.000)		-0.011 (0.000)
ATV_t		-17.53 (0.000)		-9.76 (0.000)
\bar{R}^2	[0.102]	[0.061]	[0.070]	[0.082]

Table 6: PREDICTIVE POWER OF AVERAGE VARIANCE AND CORRELATION MEASURES (NOVEMBER 1-14, 2010).

Notes: The table presents the daily averages for ordinary least squares regression results for one-step-ahead forecasting of triangular arbitrage returns (Equations 13 and 14): “Eq 13(2)” with “2” in the parentheses denotes that Equation 3 was used for the calculation of triangular returns (i.e., $r_{TR,t+1}$ from the first triangular arbitrage route) and “Eq 14(3)” with “3” in the parentheses denotes that Equation 4 was used for the calculation of triangular returns (i.e., $r_{TR,t+1}$ from the second triangular arbitrage route). The numbers in parentheses are Newey and West (1987) p-values with ten lags for the estimates and the numbers in square brackets are the \bar{R}^2 values for each predictive regression. The regressors are the average triangular variance (ATV) and correlation (ATC), defined in Equation 8 and Equation 9, respectively.

Oct. 3-16, 2011	Eq 13(2)	Eq 13(2)	Eq 14(3)	Eq 14(3)
Constant	0.999 (0.000)	0.999 (0.000)	0.999 (0.000)	0.999 (0.000)
TV_t	-9.53 (0.000)		-9.25 (0.000)	
ATC_t		-0.009 (0.000)		-0.009 (0.000)
ATV_t		-4.82 (0.000)		-5.96 (0.000)
\bar{R}^2	[0.042]	[0.048]	[0.035]	[0.036]

Table 7: PREDICTIVE POWER OF AVERAGE VARIANCE AND CORRELATION MEASURES (OCTOBER 3-16, 2011).

Notes: The table presents the daily averages for ordinary least squares regression results for one-step-ahead forecasting of triangular arbitrage returns (Equations 13 and 14): “Eq 13(2)” with “2” in the parentheses denotes that Equation 3 was used for the calculation of triangular returns (i.e., $r_{TR,t+1}$ from the first triangular arbitrage route) and “Eq 14(3)” with “3” in the parentheses denotes that Equation 4 was used for the calculation of triangular returns (i.e., $r_{TR,t+1}$ from the second triangular arbitrage route). The numbers in parentheses are Newey and West (1987) p-values with ten lags for the estimates and the numbers in square brackets are the \bar{R}^2 values for each predictive regression. The regressors are the average triangular variance (ATV) and correlation (ATC), defined in Equation 8 and Equation 9, respectively.

<u>Model</u>	<u>PERCTA</u>	<u>DM (p-value)</u>
RW	0.30	
Equation 14	0.54	10.17 (0.000)

Table 8: DIRECTIONAL FORECAST PERFORMANCE (NOVEMBER 1-14, 2010).
Notes: The table presents the daily averages for the percentage of correctly forecasted signs of the change in triangular arbitrage returns (*PERCTA*) for the random walk model (*RW*) and the model specification given by Equation 14. The Diebold and Mariano (DM) (1995) test is used to measure the statistical significance of the sign forecasts of Equation 14 over the random walk model.

To confirm the predictive power of our measures relative to the random walk model, we also perform a directional test that determines the percentage of correctly forecasted signs of the change in triangular arbitrage returns: $PERCTA = 1/T \sum_{i=1}^T \rho_t$, where $\rho_t = 1$ if $\Delta r_{TR,t+1} \widehat{\Delta r_{TR,t+1}} = 1$, and zero, otherwise. The significance of the difference in the performance of the model given by Equation 14 and the random walk model is tested by the Diebold and Mariano (1995) (DM) test statistic. The null hypothesis is that there is no difference in the percentage of correctly predicted directional movements in triangular arbitrage returns of the two alternative forecasting models. We average our findings for each day over the period from November 1-14, 2010 and use the expanding sample one-step-ahead forecasting. The DM statistic in Table 8 shows statistically significant forecast improvements over the random walk model at the 1% significance level. This exercise demonstrates the difficulties of forecasting the directional movements in triangular arbitrage returns accurately and the added value of utilizing the proposed predictors.

Next, based on our conclusions from Table 5, we run the following predictive regressions:

$$r_{TR,t+1}^i = \alpha_i + \beta_{1i} \overline{qwsprea\bar{d}}_t + \beta_{2i} r_{TRqw,t}^i + \epsilon_{i,t}, \quad i = 1, 2 \quad (27)$$

$$r_{TR,t+1}^i = \alpha_i + \beta_{1i} \overline{cogqwsprea\bar{d}}_t + \beta_{2i} r_{TRcog,t}^i + \psi_{i,t}, \quad i = 1, 2 \quad (28)$$

where $r_{TR,t+1}^i$ (triangular arbitrage returns) stands for the left-hand sides of Equation 3 ($i=1$) and Equation 4 ($i=2$), and the predictor variables are defined as follows:

- $\overline{qwsprea\bar{d}}$ is the quantity-weighted average bid-ask spread defined by Equation 18;
- $\overline{cogqwsprea\bar{d}}$ is the center of gravity quantity-weighted average bid-ask spread defined by Equation 22;
- $r_{TRqw,t}^i$ is the left-hand side of Equation 3 (when $i=1$) or Equation 4 (when $i=2$) when bid and ask prices are calculated by Equations 16 and 17;

Nov. 1-14, 2010	$\overline{qwspread}_t$	$\overline{cogqwspread}_t$	$r_{TRqw,t}^i$	$r_{TRcoq,t}^i$	\overline{R}^2
$r_{TR,t+1}^1$ ($i = 1$)					
β_{1i}, β_{2i}		-0.003		0.036	[0.212]
(p-value)		(0.000)		(0.000)	
β_{1i}, β_{2i}	-0.006		0.065		[0.263]
(p-value)	(0.000)		(0.000)		
$r_{TR,t+1}^2$ ($i = 2$)					
β_{1i}, β_{2i}		-0.006		0.016	[0.191]
(p-value)		(0.000)		(0.000)	
β_{1i}, β_{2i}	-0.010		0.047		[0.249]
(p-value)	(0.000)		(0.000)		

Table 9: PREDICTIVE POWER OF LIMIT ORDER BOOK MEASURES (NOVEMBER 1-14, 2010; DAILY AVERAGES).

Notes: $r_{TR,t+1}^i$ (triangular arbitrage returns) stands for the left-hand sides of Equation 3 ($i=1$) and Equation 4 ($i=2$). The predictors are the quantity-weighted average bid-ask spread ($\overline{qwspread}$), the center of gravity quantity-weighted average bid-ask spread ($\overline{cogqwspread}$), triangular arbitrage returns obtained from the quantity-weighted average bid and ask prices ($r_{TRqw,t}^i$), and triangular arbitrage returns obtained from the center of gravity quantity-weighted average bid and ask prices ($r_{TRcoq,t}^i$). The numbers in parentheses are Newey and West (1987) p-values (p-value) with ten lags for the estimates of β_i , and the numbers in square brackets are the \overline{R}^2 values of each predictive regression.

- $r_{TRcoq,t}^i$ is the left-hand side of Equation 3 (when $i=1$) or Equation 4 (when $i=2$) when bid and ask prices are calculated by Equations 20 and 21.

Tables 9 and 10 report the estimates of slope coefficients from Equations 27 and 28. Although all limit order book indicators are informative for predicting triangular arbitrage returns, it can be observed that the standard quantity-weighted indicators ($r_{TRqw,t}^i$) are dominant and they have the highest predictive power. The spread measures ($\overline{qwspread}$ and $\overline{cogqwspread}$) suggest that the lower the average spread, the larger future triangular arbitrage returns are expected. In particular, triangular arbitrage returns are forecasted to increase by between 10-100 bps on average when the average spread declines by one pip. Specifically, 100 bps increase in the returns on triangular arbitrage according to the limit order book measures forecasts an average increase in the actual triangular arbitrage strategy profits by 1-7 bps. This evidence demonstrates the usefulness of limit order book information to arbitrage traders. In addition, these measures are more effective in forecasting triangular arbitrage returns relative to the average covariance and correlation measures.¹⁸

¹⁸An alternative Logit model specification (where the presence of profitable arbitrage is given by an indicator variable (0=no profits, 1=profits)) yields statistically significant coefficients on all regressors (indicators), although the pseudo- R^2 values are roughly half of the \overline{R}^2 's from the original regressions. We do not report these results for brevity, but they can be available upon request. In essence, our regressions explore and predict the likelihood

Nov. 1-14, 2010	$\overline{qwsread}_t$	$\overline{cogqwsread}_t$	$r_{TRqw,t}^i$	$r_{TRcoq,t}^i$	$[\overline{R}^2]$
$r_{TR,t+1}^1 (i=1)$					
β_{1i}, β_{2i}		-0.013		0.009	[0.082]
(p-value)		(0.000)		(0.000)	
β_{1i}, β_{2i}	-0.010		0.035		[0.180]
(p-value)	(0.000)		(0.000)		
$r_{TR,t+1}^2 (i=2)$					
β_{1i}, β_{2i}		-0.002		0.039	[0.099]
(p-value)		(0.000)		(0.000)	
β_{1i}, β_{2i}	-0.006		0.042		[0.116]
(p-value)	(0.000)		(0.000)		

Table 10: PREDICTIVE POWER OF LIMIT ORDER BOOK MEASURES (OCTOBER 3-16, 2011; DAILY AVERAGES).

Notes: $r_{TR,t+1}^i$ (triangular arbitrage returns) stands for the left-hand sides of Equation 3 ($i=1$) and Equation 4 ($i=2$). The predictors are the quantity-weighted average bid-ask spread ($\overline{qwsread}$), the center of gravity quantity-weighted average bid-ask spread ($\overline{cogqwsread}$), triangular arbitrage returns obtained from the quantity-weighted average bid and ask prices ($r_{TRqw,t}^i$), and triangular arbitrage returns obtained from the center of gravity quantity-weighted average bid and ask prices ($r_{TRcoq,t}^i$). The numbers in parentheses are Newey and West (1987) p-values (p-value) with ten lags for the estimates of β_i , and the numbers in square brackets are the \overline{R}^2 values of each predictive regression.

We interpret these results by viewing the spread measures from the market liquidity cost perspective. Considering the need for trading speed in capturing triangular arbitrage, we conclude that any benefits coming from the reduction in the average spread are likely (and mostly) absorbed by high-frequency traders. Hence, decreases in the average spread benefit high-frequency traders and increase their arbitrage profits, which further raises market quality. In other words, lower average trading costs increase the size and likelihood of triangular arbitrage profits. In contrast, higher liquidity costs eliminate arbitrage opportunities, limit the role of high-frequency traders and reduce FX market quality.

In general, our findings indicate that high-frequency FX traders can profit by exploiting the shape of the order book, rather than by taking orders away from the front of the limit order book as in Cvitanic and Kirilenko (2010). Basically, Cvitanic and Kirilenko (2010) show that machine trading in an equity market extracts positive expected profits by undercutting slow human orders at the front of the book. Similar strategy in FX markets called “sub-penny jumping” is described in Mahmoodzadeh and Gençay (2016). Complementing these studies, our work uncovers the fact that triangular arbitrage will occur. We found that predictors significantly increase the probability of observing arbitrage opportunities. This type of predictive guidance is important for investors that anxiously await and search for arbitrage.

Feb. 21-27, 2011	$r_{TR,t+1}^1$	$r_{TR,t+1}^2$	Apr. 4-10, 2011	$r_{TR,t+1}^1$	$r_{TR,t+1}^2$
Freq	34		Freq	120	
ATC_t	-0.001 (0.000)	-0.002 (0.000)	ATC_t	-0.004 (0.000)	-0.008 (0.000)
ATV_t	-11.63 (0.000)	-4.62 (0.000)	ATV_t	-5.78 (0.000)	-1.13 (0.000)
\overline{R}^2	[0.032]	[0.015]	\overline{R}^2	[0.035]	[0.026]

Table 11: PREDICTIVE POWER OF ATV AND ATC FOR THE EVENT STUDY.

Notes: The table presents the daily averages for ordinary least squares regression results for one-step-ahead forecasting of triangular arbitrage returns as specified in Equation 14. $r_{TR,t+1}^i$ (triangular arbitrage returns) stands for the left-hand sides of Equation 3 ($i=1$) and Equation 4 ($i=2$). The numbers in parentheses are Newey and West (1987) p-values with ten lags for the estimates and the numbers in square brackets are the average \overline{R}^2 values for each predictive regression for the weeks Feb. 21-27, 2011 and Apr. 4-10, 2011. The regressors are the average triangular variance (ATV) and correlation (ATC), defined in Equation 8 and Equation 9, respectively. “Freq” is the average number of daily triangular parity violations over the two periods.

that FX traders may benefit from both the pace of information gathering (e.g., information related to limit book indicators) and order placement that is based on this information.

6.3. The Impact of Decimal Pip Pricing on Prediction

Decimal pip pricing refers to the addition of a fifth decimal place to the prices in the EBS platform.¹⁹ The policy was introduced to accommodate the platform’s high-frequency traders (HFT) and to respond to the potential threat from competing platforms such as the ones from Barclays and Deutsche Bank. Although the move to decimal pips accelerated a decline in the market share for EBS and the policy was subsequently scrapped in September, 2012, it is of interest to test for the impact of pip pricing on our results.

In particular, our goal here is to observe the frequency and predictability of triangular arbitrage opportunities before and after the introduction of decimal pip pricing by EBS in mid March, 2011. In the week of February 21-27, 2011 the whole pip pricing was still in use, while from April 4-10, 2011, the new decimal pip pricing was in effect. In what follows, we will apply our framework to the last week of February, 2011 and the first week of April, 2011.

We have shown before that an improved predictability of triangular arbitrage by using the ATC and ATV predictors implies forecast gains from the order book indicators. For consistency now, we hence employ only ATC and ATV as regressors. Table 11 presents the daily average estimates from

¹⁹In the context of basis points, considering that currencies are typically quoted to four decimal places, one pip corresponds to one basis point.

our predictive regressions. The average daily return from triangular arbitrage in the week of Feb. 21-27, 2011 is 0.34 bps, while this figure for the week of Apr. 4-10, 2011 is 0.46 bps. It is worthwhile to mention that both figures are lower than the averages for Nov. 1-14, 2010 (0.63 bps) and Oct. 3-16, 2011 (0.59 bps). A somewhat surprising finding is that the number of average triangular arbitrage opportunities plunges to about 34 before the structural change and then increases to very high levels (120, on average) after the change. In addition, we observe that both trading volume and trading frequency were lower than average before the regulation. This can be explained by the potential behavior of market participants that may have pulled back due to uncertainty to absorb the structural change. This also caused the reduced predictability of triangular arbitrage opportunities.

After the change to decimal pip pricing, the new platform setting attracted HFT and this resulted in an increase in trading volume, number of arbitrage opportunities, average profitability and a better regression fit as measured by the \overline{R}^2 . Following that, the average predictability and profitability slightly improved later in 2011, but the number of triangular arbitrage situations fell below the 2010 levels (to roughly 80 in October, 2011, which is the last available month in our data set). According to Reuters, the average daily cash FX volume on the EBS platform dropped by 49% from August 2011 to August 2012, when it was \$95.5 billion. This decline in trading activity, likely caused by the departure of traders and banks that used slower technology relative to HFT, is consistent with our results.²⁰

7. Conclusion

Triangular arbitrage strategy involves exploiting mispricing in the FX market when a currency is traded at two different prices. Prior research documents that triangular arbitrage situations are often difficult to profit from due to delays in trade execution, technological advances and competition among traders. The conventional view is that these features restrict the timing and

²⁰Of course, our focus in this paper is on understanding the triangular arbitrage for actively traded currencies and thus it is particularly relevant for major FX rates. Nevertheless, from a methodological perspective, our approach is considerably generic, and it can be applied to other currencies at ease. To achieve a more global stand, one specific idea (as we leave it for future research) would be to combine our wavelet-based setup with the theoretical framework of Cui *et al.* (2018), who generalize triangular currency arbitrage to “*k*-currency” arbitrage. From the empirical viewpoint, however, such a multi-currency/country assessment requires significant caution as well as careful data processing. This is mainly due to the lack of high quality data availability at the “ultra”-high sampling frequencies. The currencies that we considered in our study are the major currencies for which the market is liquid and data quality is very high accordingly. Regardless, an empirical extension would further consider the G-10 currency pairs (see e.g., Cui *et al.*, 2018) as long as data quality is ensured and sampling frequencies are set high. The noise component or false detection of arbitrage is, however, likely to occur when other currencies are considered, particularly, those that are not liquid enough, such as emerging market currencies. We, therefore, suggest applied researchers to consider major currencies to study triangular arbitrage through our modeling choice.

prediction of arbitrage opportunities.

This paper introduces an alternative framework that contributes to our understanding and predicting triangular arbitrage. We first propose using a robust prediction method that relies on wavelets in a regression setup. Unlike the existing techniques, our approach is considerably immune to noisy data and measurement biases which likely create misidentification of arbitrage. We describe this potential statistical problem and propose a solution based on a wavelet-based estimation. We then demonstrate empirically the existence and properties of triangular arbitrage in three major exchange rates. Although the observed short durations indicate that markets are efficient in terms of limiting the trade gain, we find evidence that the use of *ultra*-high-frequency data within our framework allows investors to exploit arbitrage profits.

The quality of our unique data coming from the EBS platform enables us to go one step further and identify the variables that may be used to predict triangular arbitrage profits. We find that two sets of predictors are statistically and economically informative: 1) average variance and average correlation among the exchange rates, and 2) limit order book indicators averaged across the three exchange rates from the triangular parity condition. In sharp contrast to other studies, our original limit order book measures are based on the quantity-weighted average bid/ask prices and the center of gravity quantity-weighted average bid/ask prices. These measures are dominant in predicting arbitrage profits relative to the average variance and correlation measures. Specifically, we show that when the average FX volatility and average correlations among the FX rates are low, triangular arbitrage returns are expected to increase. Moreover, the forecast accuracy significantly improves as we augment wavelet-based predictive regressions by including market microstructure measures.

Deviations from FX parity conditions offer distinct opportunities to financial managers. For instance, managers can make profit-maximizing decisions by exploiting deviations from the parity conditions over certain periods of time, or they may wish to avoid or hedge the risks of such deviations. If the triangular parity is violated, it is essential for the analyst to know the time-series properties of deviations. Moreover, if deviations can be predicted, then speculative strategies can be profitable. Specifically, when the average deviation is non-zero and volatility is low, our results suggest that managers may be attracted to a speculative strategy. But, if deviations have a relatively high volatility, managers would need to weigh the risk-return tradeoff. Another interpretation of this phenomenon is that investors may use limit orders when the market is stable. When the market is volatile, investors use market orders to trade quickly. Considering that price impacts of market orders are generally larger than price impacts of limit orders, market orders may eliminate triangular arbitrage quicker and diminish the usefulness of the limit order book in arbitrage.²¹ Finally, we find it important to note that central banks and government agencies

²¹We are grateful to the three anonymous referees for this and other useful suggestions.

also trade currencies as part of their economic and financial policy. Hence, our framework could potentially provide information on how closely national financial markets are linked to one another via exchange rates, which can be extended to broader implications for financial stability, financial regulation and government intervention (e.g., coordinated central bank interventions).

Our research can be extended in several important lines. First, based on the predictive regression we develop here, it is possible to link the precise timing of realized triangular arbitrage opportunities to shocks driven by macro news announcements or monetary policy decisions. This implementation is likely to increase the forecasting power of exogenous variables. From a broader perspective, the second research path would be to explore the conditional predictability of triangular arbitrage also for other exchange rates, primarily the Swiss franc and the British pound. While we mainly examine the role of economic uncertainty in arbitrage prediction, political uncertainty—as an alternative channel—could also help explain arbitrage dynamics (see e.g., Kelly *et al.*, 2016 on the link between political uncertainty and option prices). Fourth, on the methodological front, our arbitrage measurement device can incorporate not only volatility risk and correlation, but also realized price jumps in line with Bajgrowicz *et al.* (2016). Jumps convey predictive information for FX volatility, which may improve the accuracy of arbitrage predictions. Lastly, one can utilize predictors other than the measures of liquidity. Studying the impact of credit risk and the order size on triangular arbitrage returns might provide new insights into the feasibility of arbitrage profits.

Appendix

Our wavelet-based (robust) least square estimation relies on discrete wavelet transform (DWT) and corresponding variance analysis. This appendix details the statistical properties and characteristics behind these techniques. We refer the reader to Gençay and Gradojevic (2011) for a general discussion on wavelet framework.

In principle, wavelet analysis can be carried out in all arbitrary time scales. This may not be necessary if only key features of the data are in question, and if so, DWT is an efficient and parsimonious route as compared to the continuous wavelet transformation (CWT). The DWT is a subsampling of $w(\lambda, t)$ with only dyadic scales, i.e., λ is of the form 2^{j-1} , $j = 1, 2, 3, \dots$ and, within a given dyadic scale 2^{j-1} , t 's are separated by multiples of 2^j .

Let \mathbf{x} be a dyadic length vector ($N = 2^J$) of observations. The length N vector of discrete wavelet coefficients \mathbf{w} is obtained by

$$\mathbf{w} = \mathcal{W}\mathbf{x},$$

where \mathcal{W} is an $N \times N$ real-valued orthonormal matrix (based on the wavelet type) defining the DWT which satisfies $\mathcal{W}^T \mathcal{W} = I_N$ ($n \times n$ identity matrix).²² The n th wavelet coefficient w_n is associated with a particular scale and with a particular set of times. The vector of wavelet coefficients may be organized into $J + 1$ vectors,

$$\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_J, \mathbf{v}_J]^T,$$

where \mathbf{w}_j is a length $N/2^j$ vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$ and \mathbf{v}_J is a length $N/2^J$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$.

Using the DWT, we may formulate an additive decomposition of \mathbf{x} by reconstructing the wavelet coefficients at each scale independently. Let $\mathbf{d}_j = \mathcal{W}_j^T \mathbf{w}_j$ define the j th level *wavelet detail* associated with changes in \mathbf{x} at the scale λ_j (for $j = 1, \dots, J$). The wavelet coefficients $\mathbf{w}_j = \mathcal{W}_j \mathbf{x}$ represent the portion of the wavelet analysis (decomposition) attributable to scale λ_j , while $\mathcal{W}_j^T \mathbf{w}_j$ is the portion of the wavelet synthesis (reconstruction) attributable to scale λ_j . For a length $N = 2^J$ vector of observations, the vector \mathbf{d}_{J+1} is equal to the sample mean of the observations.

A multiresolution analysis (MRA) may now be defined via

$$x_t = \sum_{j=1}^{J+1} \mathbf{d}_{j,t} \quad t = 1, \dots, N. \quad (29)$$

That is, each observation x_t is a linear combination of wavelet detail coefficients at time t . Let $\mathbf{s}_j = \sum_{k=j+1}^{J+1} \mathbf{d}_k$ define the j th level *wavelet smooth*. Whereas the wavelet detail \mathbf{d}_j is associated with variations at a particular scale, \mathbf{s}_j is a cumulative sum of these variations and will be smoother and smoother as j increases. In fact, $\mathbf{x} - \mathbf{s}_j = \sum_{k=1}^j \mathbf{d}_k$ so that only lower-scale details (high-frequency features) from the original series remain. The j th level *wavelet rough* characterizes the remaining lower-scale details through

$$\mathbf{r}_j = \sum_{k=1}^j \mathbf{d}_k, \quad 1 \leq j \leq J + 1.$$

The wavelet rough \mathbf{r}_j is what remains after removing the wavelet smooth from the vector of observations. A vector of observations may thus be decomposed through a wavelet smooth and rough via

$$\mathbf{x} = \mathbf{s}_j + \mathbf{r}_j,$$

²²Since DWT is an orthonormal transform, orthonormality implies that $\mathbf{x} = \mathcal{W}^T \mathbf{w}$ and $\|\mathbf{w}\|^2 = \|\mathbf{x}\|^2$.

for all j .

The terminology “detail” and “smooth” were used by Percival and T. (2000) to describe additive decompositions from Fourier and wavelet transforms. The goal is to look at data at different resolutions with this representation. The smooth part is coarse: we are looking at local averages, i.e., low-frequency trends and the sample mean. The detail is deviation from the smooth part.

A variation of the DWT is called the maximum overlap DWT (MODWT). Similar to the DWT, the MODWT is a subsampling at dyadic scales, but in contrast to the DWT, the analysis involves all times t rather than the multiples of 2^j . Retainment of all possible times eliminates alignment effects of DWT and leads to more efficient time series representation at multiple time scales.

References

- Aiba, Y. and Hatanoa, N. (2004). Triangular arbitrage in the foreign exchange market. *Physica A*, **344**, 174–177.
- Aloosh, A. and Bekaert, G. (2019). Currency factors. Columbia Business School Research Paper No. 17-18.
- Bajgrowicz, O., Scaillet, O., and Treccani, A. (2016). Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science*, **62**(8), 2198–2217.
- Bali, T. G. (2007). Modeling the dynamics of interest rate volatility with skewed fat-tailed distributions. *Annals of Operations Research*, **151**, 151–178.
- Bjønnes, G. and Longarela, I. (2014). Arbitrage violations in currency markets. Working Paper.
- Bloomfield, R., O’Hara, M., and Saar, G. (2005). The “make or take” decision in an electronic market: Evidence on the evolution of liquidity. *Journal of Financial Economics*, **75**(1), 165–199.
- Cao, C., Hansch, O., and Wang, X. (2009). The information content of an open limit-order book. *Journal of Futures Markets*, **29**(1), 16–41.
- Caporin, M., Ranaldo, A., and Velo, G. G. (2015). Precious metals under the microscope: a high-frequency analysis. *Quantitative Finance*, **15**(5), 743–759.
- Cenedese, G., Sarno, L., and Tsiakas, I. (2014). Foreign exchange risk and the predictability of carry trade returns. *Journal of Banking and Finance*, **42**, 302–313.
- Chaboud, A., Chiquoine, B., Hjalmarsson, E., and Vega, C. (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *Journal of Finance*, **69**(5), 2045–2084.
- Choi, M. S. (2011). Momentary exchange rate locked in a triangular mechanism of international currency. *Applied Economics*, **43**(16), 2079–2087.
- Christoffersen, P. F. and Diebold, F. X. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, **52**(8), 1273–1287.
- Cui, Z., Qian, W., Taylor, S. M., and Zhu, L. (2018). Detecting arbitrage in the foreign exchange market. Stevens Institute of Technology School of Business Research Paper.
- Cvitanic, J. and Kirilenko, A. A. (2010). High frequency traders and asset prices. Working Paper.

- Daniélsson, J., Luo, J., and Payne, R. (2011). Exchange rate determination and intermarket order flow effects. *European Journal of Finance*, **18**(9), 823–840.
- Darolles, S., Fol, G. L., and Mero, G. (2017). Mixture of distribution hypothesis: Analyzing daily liquidity frictions and information flows. *Journal of Econometrics*, **201**, 367–383.
- Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263.
- Evans, M. D. D. and Lyons, R. K. (2012). Exchange rate fundamentals and order flow. *Quarterly Journal of Finance*, **2**(4), 1–63.
- Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, **102**(480), 1349–1362.
- Fenn, D. J., Howison, S. D., McDonald, M., Williams, S., and Johnson, N. F. (2009). The mirage of triangular arbitrage in the spot foreign exchange market. *International Journal of Theoretical and Applied Finance*, **12**(8), 1105–1123.
- Foucault, T., Kozhan, R., and Tham, W. (2017). Toxic arbitrage. *Review of Financial Studies*, **30**(4), 1053–1094.
- Gençay, R. and Gradojevic, N. (2011). Errors-in-variables estimation with wavelets. *Journal of Statistical Computation and Simulation*, **81**(11), 1545–1564.
- Gençay, R. and Gradojevic, N. (2013). Private information and its origins in an electronic foreign exchange market. *Economic Modelling*, **33**, 86–93.
- Giannellis, N. and Papadopoulos, A. P. (2009). Testing for efficiency in selected developing foreign exchange markets: An equilibrium-based approach. *Economic Modelling*, **26**(1), 155–166.
- Gould, M. D., Porter, M. A., and Howison, S. D. (2017). Quasi-centralized limit order books. *Quantitative Finance*, **17**(6), 831–853.
- Harris, L. and Panchapagesan, V. (2005). The information content of the limit order book: Evidence from NYSE specialist trading decisions. *Journal of Financial Markets*, **8**(1), 25–67.
- Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *Journal of Finance*, **50**, 1175–1199.
- Haven, E., Liu, X., and Shen, L. (2012). De-noising option prices with the wavelet method. *European Journal of Operational Research*, **222**(1), 104–112.

- Ito, T., Yamada, K., Takayasu, M., and Takayasu, H. (2012). Free lunch! Arbitrage opportunities in the foreign exchange markets. NBER Working Paper No. 18541.
- Kaniel, R. and Liu, H. (2006). So what orders do informed traders use? *Journal of Business*, **79**(4), 1867–1913.
- Kelly, B., Lubos, P., and Veronesi, P. (2016). The price of political uncertainty: Theory and evidence from the option market. *Journal of Finance*, **71**(5), 2417–2480.
- Knif, J., Pynnönen, S., and Luoma, M. (1995). An analysis of lead-lag structures using a frequency domain approach: Empirical evidence from the Finnish and Swedish stock markets. *European Journal of Operational Research*, **81**(2), 259–270.
- Kozhan, R. and Salmon, M. (2012). The information content of a limit order book: The case of an FX market. *Journal of Financial Markets*, **15**(1), 1–28.
- Kozhan, R. and Tham, W. W. (2012). Execution risk in high-frequency arbitrage. *Management Science*, **58**(11), 2131–2149.
- Lyons, R. K. and Moore, M. J. (2009). An information approach to international currencies. *Journal of International Economics*, **79**(2), 211–221.
- Mahmoodzadeh, S. and Gençay, R. (2016). Human vs. high frequency traders in the interbank FX market, role of tick size. Working Paper.
- Moore, M. J. and Payne, R. (2011). On the sources of private information in FX markets. *Journal of Banking and Finance*, **35**, 1250–1262.
- Mueller, P., Tahbaz-Salehi, A., and Vedolin, A. (2017). Exchange rates and monetary policy uncertainty. *The Journal of Finance*, **72**, 1213–1252.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.
- Pasquariello, P. (2014). Financial market dislocations. *Review of Financial Studies*, **27**(6), 1868–1914.
- Percival, D. B. and T., W. A. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge Press, Cambridge.
- Piccotti, L. R. (2018). Jumps, cojumps, and efficiency in the spot foreign exchange market. *Journal of Banking and Finance*, **87**, 49–67.

- Schaumburg, E. (2014). Has automated trading promoted efficiency in the FX spot market? Federal Reserve Bank of New York: Liberty Street Economics Blog.
- Sun, E. W. and Meinel, T. (2012). A new wavelet-based denoising algorithm for high-frequency financial data mining. *European Journal of Operational Research*, **217**(3), 589–599.
- Sun, E. W., Chen, Y.-T., and Yu, M.-T. (2015). Generalized optimal wavelet decomposing algorithm for big financial data. *International Journal of Production Economics*, **165**, 194–214.
- Wu, J.-L. and Chen, P.-F. (2008). A revisit on dissecting the PPP puzzle: Evidence from a nonlinear approach. *Economic Modelling*, **25**(4), 684–695.