Currency Trading in Bitcoin Markets

A Study of the USD/EUR Pair

ビットコイン市場における通貨取引

－米ドル/ユーロペアに関する研究－

A Dissertation Presented to
the Graduate School of Arts and Sciences
International Christian University
for the Degree of Doctor of Philosophy

国際基督教大学　大学院
アーツ・サイエンス研究科提出博士論文

December 6, 2020

NAN, Zheng
南　峥

Currency Trading in Bitcoin Markets:
A Study of the USD/EUR Pair

ビットコイン市場における通貨取引

－米ドル/ユーロペアに関する研究－

A Dissertation Presented to
the Graduate School of Arts and Sciences
International Christian University
for the Degree of Doctor of Philosophy

国際基督教大学　大学院
アーツ・サイエンス研究科提出博士論文

December 6, 2020
２０２０年１２月６日

NAN, Zheng
南　峥

審査委員会メンバー

Members of Evaluation Committee

主査 / Chief Examiner

　海蔵寺 大成　教授

副査 / Examiner

　金子 拓也　准教授

副査 / Examiner

　MONTGOMERY, Heather A.　教授

Acknowledgements

Five years of study (two years in the master's degree program) in the Graduate School of Arts and Sciences at the International Christian University have contributed to this thesis. I thank my advisor, Prof. Taisei Kaizoji, for his visionary guidance and continuous support for five years, generously, patiently, and kindly. Because of his recommendation, I have had the chance to study in the Department of Mathematics at the University of Sussex as a visiting researcher for two months.

I also wish to thank Prof. Enrico Scalas from the University of Sussex for his wisdom, humor, kindness, and company. He, Mrs. Scalas, and his excellent colleagues helped and cared for me when I was in the U.K.

I am grateful to Prof. Heather Montgomery and Prof. Takuya Kaneko for their valuable comments and suggestions that resulted in improvements in my thesis. I thank Prof. Jun Saito for giving me a foreign language test.

I would also like to thank Prof. Insang Hwang and Prof. Tomo Inoue for giving me a solid foundation in econometrics and thank Prof. Lukas Pichl for teaching me machine learning and his collaborations. I thank the teachers from the Japanese Language Program office who encouraged me to learn Japanese.

I want to express my gratitude to the Graduate School office for their help and tremendous support. I wish to thank the Center for Research Planning and Support of ICU for the enormous time and energy they devoted to helping me obtain research funding.

I am grateful to the Reference Support Center of the library for giving me a part-time job to support my family for many years.

Thanks also to my friends Hiroaki Watanabe, Nao Satoh, Eiman Fouad, William Wen, and Cleverson Souza for their kindness, care, and prayers.

I want to express my gratitude to the Mitsubishi Scholarship for international students and the ICU Doctoral Researcher Funds, both type-A and type B, for funding my study and life.

Finally, I want to thank my parents for their unfailing support, and my dear wife, Qiu Wang, for her selfless support and loving care for our family, even as she is embarking on her Ph. D program. I want to say, “Thank you” and “I love you” to my wife, my daughter, Xiang Nan, and my son, Ding Nan. I enjoy every day that we spend together.

Publications

[1] Nan, Z., & Kaizoji, T. (2019a). Bitcoin-based triangular arbitrage with the Euro/US dollar as a foreign futures hedge: modeling with a bivariate GARCH model. *Quantitative Finance and Economics*, 3(2), 347.

[2] Nan, Z., & Kaizoji, T. (2019b). Market efficiency of the bitcoin exchange rate: Weak and semi-strong form tests with the spot, futures and forward foreign exchange rates. *International Review of Financial Analysis*, 64, 273-281.

[3] Nan, Z., & Kaizoji, T. (2020). The Optimal Foreign Exchange Futures Hedge on the Bitcoin Exchange Rate: An Application to the US Dollar and the Euro. In *Advanced Studies of Financial Technologies and Cryptocurrency Markets* (pp. 163-181). Springer, Singapore.

[4] Pichl, L., Nan, Z., & Kaizoji, T. (2020). Time series analysis of ether cryptocurrency prices: Efficiency, predictability, and arbitrage on exchange rates. In *Advanced studies of financial technologies and cryptocurrency markets* (pp. 183-196). Springer, Singapore.

Table of Contents

1 Introduction

As worldwide bitcoin transactions occur every day, bitcoin markets perform like a new form of "foreign exchange markets" in many ways.

Two domestic bitcoin exchanges that trade bitcoins for different currencies could together play an implicit role as foreign exchange. Consider a simple case: After trading three U.S. dollars for a bitcoin in a U.S. bitcoin market, an American user can sell the bitcoin for two Euros in a European bitcoin market. These trades form a USD/EUR price of 3/2 = 1.5, indicating the U.S. dollar price of a Euro in the bitcoin markets. Exchanging U.S. dollars into Euros is performed simply by trading with bitcoins twice.

Some major bitcoin exchanges that have developed into multinationals also facilitate bitcoin-based currency trading. Coinbase, a San Francisco-based bitcoin exchange, has expanded to over 100 countries and regions. Transferring bitcoins from account to account within an international brokerage appears to be an efficient option and has lower costs than doing so across exchanges. Different currencies are hence exchangeable with the help of bitcoins.

Some FX brokers, such as Ava Trade, eToro, and LiteForex, have begun to accept bitcoins for currency trading. Though they may aim to add a more dynamic dimension to forex trading, these brokers make bitcoins that can be traded for several different currencies using a single brokage account –further facilitating the proposed bitcoin-based forex trade.

The boundary that discriminates a bitcoin exchange from a foreign exchange thus becomes ambiguous.

Bitcoin markets may well be an appealing alternative to foreign exchange markets. Here, the term "bitcoin markets" refers to a broad concept covering any marketplace that trades bitcoins. Under this assumption, learning the behavior of foreign exchange prices in the bitcoin markets becomes crucial to appropriate modeling and accurate forecasting.

This chapter begins with my motivation for choosing bitcoin-based USD/EUR trading as my research topic, followed by my research objectives, methodology, and a guide to each chapter.

1.1 Motivations

Bitcoin trading has a number of fascinating features, including decentralized valuations, low costs of trading, no geographical boundaries, low deposits, and high leverage. These features could benefit the bitcoin-based forex trade. Even conventional currency trading may change, at least in some ways, after the forex brokers incorporate bitcoins into their business.

The problem is how to identify the bitcoin-based foreign exchange trade from daily bitcoin trading and then how to measure it. Central to this problem is finding the exchange rate between two fiat currencies in bitcoin markets.

Nan & Kaizoji (2017, 2019b) define the instantaneous price of Euros in U.S. dollars in a bitcoin market as

$$The\ USD/EUR\ bitcoin\ exchange\ rate = \frac{USD/BTC}{EUR/BTC} = (USD/EUR)^{BX} \quad (1)$$

where USD/BTC and EUR/BTC are the prices of bitcoins in U.S. dollars and Euros, respectively.

The term bitcoin exchange rate (or the BX rate) is used to distinguish the bitcoin-based exchange rate from the foreign exchange rate (or the FX rate). The BX superscript used here is intended to clearly indicate the distinction. Note that, while the bitcoin exchange rate may refer to bitcoin prices in the literature, in this thesis it is used only as the bitcoin-based foreign exchange rate.

The USD/EUR bitcoin exchange rate has captured my great interest and motivated my effort to uncover the mystery behind it and its implications for bitcoin-based currency trading. Specifically, this thesis focuses on the following research questions:

(i) What are the characteristics of the bitcoin exchange rate?

(ii) How can we model the behavior of the bitcoin exchange rate?

(iii) How can we predict the future evolution of the bitcoin exchange rate?

1.2 Research objectives

The bitcoin-based foreign exchange rate, defined as the bitcoin exchange rate in Nan & Kaizoji (2017, 2019b), offers a pivotal clue to answering the research questions listed in Section 1.1. This thesis focuses on characterizing, modeling and forecasting the USD/EUR bitcoin exchange rate behavior. These three elements can be described as follows:

(1) Characterizing – The aim is to give a thorough report of the BX rate series containing statistical descriptions and the stylized facts summarized from scientific approaches. The intent is to obtain a "global" view of the bitcoin exchange rate with little prior knowledge.

(2) Modeling – The goal is to capture the temporal evolution of the BX rate series from different perspectives, precisely, from the long-run and short-run views. The long-run equilibrium relationship between the BX rate and the FX spot rate is of interest

because the law of one price indicates the same exchange rate. It is also essential to examine whether the prevailing economic theories also apply to the bitcoin exchange rate. In the short run, the adjustment process to equilibrium is investigated.

(3) Forecasting – Past observations of the BX rate series are used to predict the future state. Outperforming the forecasting baseline is the critical factor in measuring the accuracy of the state-of-the-art approaches.

1.3 Methodology

This thesis applies various quantitative approaches to time series data. A variety of methods involving statistical tests, econometrical models, and machine learning algorithms are used. Specifically, the contents of the sections and subsections of the thesis can be summarized as follows:

- Section 3.2: Statistical descriptions of the level data series and the return series.
- Section 3.3.1: Random walk hypothesis, weak-form efficient market hypothesis, the ADF test, KPSS test, ERS test, and Ljung-Box test.
- Section 3.3.2: ARMA (p, q) model and the ARMA [p_i, q_j] model.
- Section 3.3.3: ARMA (p, q) plus GARCH (1, 1) model with the normal distribution, student's t distribution, and the GED, respectively; volatility comparisons with the static standard deviation, 150-day moving average standard deviation, and the conditional standard deviation.
- Section 3.3.4: Comparison of Value-at-Risk (VaR): the historical VaR, the mean-modified VaR, the conditional VaR from the ARMA (p, q) plus GARCH (1, 1) model.
- Section 3.3.5: Extended Chow test – three-step grid searching for two breaks; Bai-Perron test using the threshold autoregressive (TAR) model for five breaks.

- Section 4.1.1: Engle and Granger methodology using the OLS method to capture the long-run equilibrium relationship.
- Section 4.1.2: Application of unrestrictive VAR model to the Johansen tests to test for cointegration; long-run equilibrium relation using cointegrating vector $\boldsymbol{\beta}$.
- Section 4.2: Inference on $\boldsymbol{\beta}$ using the maximum likelihood ratio test; unbiased estimator hypothesis, indicative of semi-strong form market efficiency.
- Section 4.3: Covered interest parity on both the bitcoin exchange rate and the FX spot rate using the Johansen test.
- Section 4.4: Proposal of the non-linear TAR model using the previous triangular arbitrage series as the threshold to find the attractor and to show the asymmetric adjustment process.
- Section 5.1: VECM representations of the estimated Johansen models and inference on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.
- Section 5.2: Visualization of the impulse response function (IRF) to present the responses of the variables to shocks.
- Section 5.3: Momentum-threshold autoregressive (M-TAR) model to look at the non-linear effect depending on whether the previous deviation increases or decreases.
- Section 5.4: Comparison of the nonlinear ECM model with the linear VECM model.
- Section 6.1: Linear specification of the bitcoin exchange rate and its FX futures hedge.
- Section 6.2: Optimal ratio model that maximizes the user's mean-variance utility.
- Section 6.3: VECM plus DCC-GARCH (1, 1) model to capture the joint conditional distribution of the returns of the bitcoin exchange rate and the FX spot rate.
- Section 6.5: Rebalancing strategy that incorporates transaction costs.

- Section 6.6: Triangular arbitrage strategy with the FX futures hedge, which can be modeled through the ARMA plus DCC-GARCH framework.
- Section 7.1: Example of volatility forecasting using the resulting VECM plus DCC-GARCH framework.
- Section 7.2: Use of a densely connected neural network and gated recurrent units (GRU) neural network with 5-minute data to challenge the one-day-ahead bitcoin exchange rate forecast using the random walk model.

1.4 Significance

This thesis gives a global view of U.S. dollar and Euro trading in bitcoin markets using several statistical tests, econometric time-series models, and machine learning algorithms. It shows that the USD/EUR bitcoin exchange rate is the best estimator relative to the other USD/EUR prices in the bitcoin market that consider the confirmation time and the bid-ask spread.

Critical economic hypotheses such as the random walk hypothesis, the efficient market hypothesis, the law of one price, and covered interest rate parity that may never have been be applied to the bitcoin market are investigated. The results show that they all reconcile with the bitcoin markets, with some conditions.

The proposed separation of three different regimes regarding the bitcoin exchange rate provides a convenient way to consider the evolution of the bitcoin markets in terms of currency trading.

This thesis proposes various approaches and methodologies that can be applied to the bitcoin exchange rate series and other series. In addition to linear specifications, nonlinear models are also examined. As an illustration, this thesis proposes two close real-

world trading strategies with FX futures as the hedge. The proposed GRU neural network is shown to significantly out-perform the random walk model in forecasting.

1.5 Structure of the thesis

The remainder of the thesis is organized as follows: Chapter 2 reviews the literature; Chapter 3 characterizes the data series; Chapter 4 models the long-run equilibria; Chapter 5 presents the short-run dynamics; Chapter 6 illustrates two trading strategies; Chapter 7 focuses on forecasting.

2 Literature review

Bitcoin is the first widely used cryptocurrency (Bloomberg). In its peer-to-peer network, bitcoin can be directly sent from one party to another without the need for a trusted, central intermediary (Commodity Futures Trading Commission, 2017).

Decentralization is a striking feature that distinguishes cryptocurrencies such as bitcoin from E-commerce payments. Having a central authority that works as a trusted third party "increases the transaction cost, limiting the minimum practical transaction size and cutting off possible small casual transactions," as Satoshi Nakamoto (2008), a pseudonym used by the individual or group claiming to have created bitcoin, wrote in a white paper.

There are a number of strengths associated with bitcoin transactions:

- "Merchants accepting bitcoins pay fees that range from zero to less than 1%," much lower than the 2-3% typically imposed by credit card processors (Wewege & Thomsett, 2019, p. 43).
- The degree of bitcoin's divisibility is much greater than most fiat currencies; A Satoshi, the smallest unit, is equal to one hundred millionth of a bitcoin, while a dollar is divisible only into 100 cents (Kelleher, 2020).
- Bitcoins can be simultaneously transferred to multiple targets, and the transaction size does not affect the transaction fee (Nakamoto, 2008).

Proponents believe that bitcoin represents both an innovation in the established money system, which is centrally coordinated and less transparent, and a new form of currency (Bjerg, 2016; Böhme et al., 2015; Gün, 2014).

On the other hand, without the supervision of any central authority, bitcoin faces issues of money laundering and the trafficking of illegal substances of various kinds (Brezo & Bringas, 2012). Critics argue that it is nothing but a Ponzi scheme soon to collapse (as cited in Bjerg, 2016).

2.1 Bitcoin transfer and its fees and confirmation time

Bitcoins can be directly transferred from one bitcoin wallet to another. However, bitcoins do not physically exist in any bitcoin wallet. A bitcoin is an encrypted record that contains all the historical transfer information. As explained in Böhme et al. (2015), the record is stored in bitcoin's peer-to-peer network, widely replicated and publicly verifiable; any user in the network can verify the ownership of the bitcoin by checking the last transfer (specifically, the sender and the receiver) in the record, and, recursively, the history of the transfers can be traced back. A wallet is simply a node in the network that provides access to the transfer recode. By design, a bitcoin transfer is free to the users, as the transfer of a bitcoin is done by adding a piece of information to the record. A more detailed explanation of the underlying technology of bitcoin transfer is given in Böhme et al. (2015).

In practice, there are some fees and costs associated with bitcoin transfers. Firstly, new bitcoins become a reward for the effort of verifying the blockchain. In bitcoin's peer-to-peer network, there is a need for some users to voluntarily process and verify bitcoin transfers. These users—called "miners"—play a critical role in the bitcoin realm. When a transfer is listed on the bitcoin network awaiting confirmation, miners are working on solving a mathematical puzzle based on the pre-existing transfer records to verify the new

transfer. They then pack the recent transfers into a group, called a block, and after six confirmations from other miners, the confirmed block is added to the sequence of blocks, namely, the blockchain (Böhme et al., 2015). The miner who adds the newly confirmed block to the blockchain is rewarded with bitcoins that are newly issued. In a sense, the new bitcoins are viewed as a cost for miners to provide their service.

Secondly, the miners' efforts to update the blockchain carry high costs. Böhme et al. (2015) point out that the computer calculations "consume more than 173 megawatts of electricity continuously," approximately equal to 20 percent of an average nuclear power plant.

Thirdly, some users pay additional fees to prioritize the transfer queue, reducing the wait times. The fees are expressed in Satoshis per byte, where a Satoshi is one hundred millionth of a bitcoin, and a transfer is usually above 200 bytes (Buchko, 2017).

The confirmation time also varies from transfer to transfer. It usually takes 10 minutes for a miner to verify a block, and a block needs six confirmations before it is finally linked to the consensus blockchain, so an average transfer could be expected to take approximately one hour. Buchko (2017) found that when the network was crowded at the end of 2017, the average confirmation time actually exceeded 16 hours. The data provided by blockchain.com show that the average time for a transfer with miner fees ranged from about 6 minutes to over 339 minutes (roughly 5.6 hours) over the last year, while the median time for a transfer with miner fees varied from roughly 3 minutes to 19 minutes (see Figure 1).

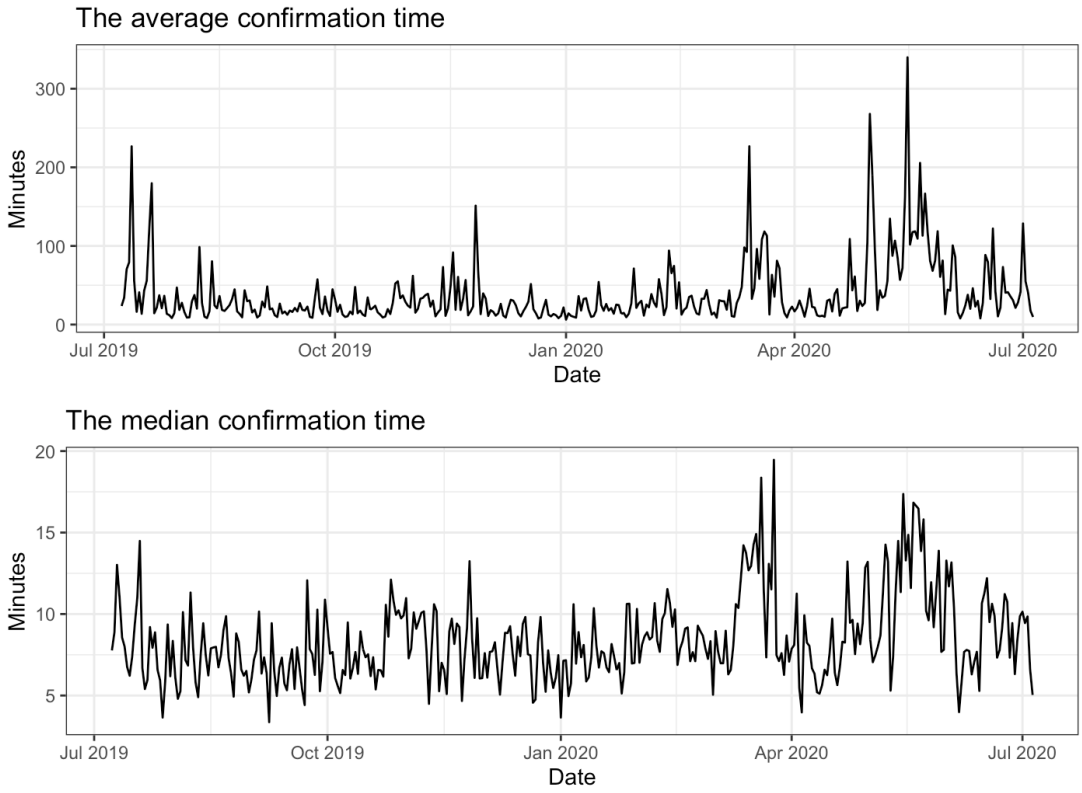

Figure 1. The average conformation time and the median confirmation time for a transfer with miner fees over one year (blockchain.com).

Bitcoin transfer speed depends on many factors. Two factors are especially critical: congestion in the bitcoin network and bitcoin transfer fees (Buchko, 2017). Easley et al. (2019) found that equilibrium transfer fees evolve in the bitcoin ecology. In the low-transfer regime, transfers without fees are posted to the blockchain, while in the high-transfer regime, only transfers with fees attached are written to the ledger. This study also found that before the issuing of new bitcoin reaches its limit, bitcoin transfer fees play only a secondary role in evoking the miners' willingness to participate and are less significant than the mining reward.

2.2 Bitcoin as money

Bjerg (2016) posits that bitcoin is commodity money without gold, fiat money without a state, and credit money without debt. In contrast, Yermack (2013) and Ali et al. (2014) argue that if a cryptocurrency is considered money, it needs to serve the three functions of money, acting as

- a store of value,
- a medium of exchange, and
- a unit of account.

Additionally, Kelleher (2020) considers the other six factors that bitcoin needs to qualify as money: scarcity, divisibility, utility, transportability, durability, and difficulty counterfeiting.

Yermack (2013) argues that bitcoin performs poorly with regard to the three criteria, in that widely swinging bitcoin prices undermine its role as a store of value, the extremely limited acceptance of bitcoins in daily commerce erodes its function as a medium, and the price of goods quoted to several decimal places discourages use of the bitcoin as a unit of account. The fact that bitcoin is still not broadly used in the conduct of retail business calls into question bitcoin's role as money (Hong, 2017).

On the other hand, Kelleher (2020) argues that bitcoin outperforms the fiat currencies in meeting the six qualifications noted above. Bitcoin is said to be a stable unit of account due to the predetermined total number of bitcoins (21 million), which helps reduce the user's menu costs (Dong & Dong, 2015).

2.3 Bitcoin as an investment tool

Individuals dealing in digital currencies are primarily interested in an alternative investment rather than an alternative transaction system (Glaser et al., 2014). As a unique investment asset, bitcoin has occupied the first market place among all cryptocurrencies, partially due to its first-mover advantage (Luther, 2016). Before 2014, bitcoin investors preferred to adopt the buy-and-hold strategy, regardless of the arbitrage opportunity (Dong & Dong, 2015). Bitcoin was soon seen by many as a unique investment asset, with very high volatility and potentially high returns, and a low correlation with other assets (Briere et al., 2013). Bouri et al. (2017) found that bitcoin could generally serve as a useful diversifier because of its weak positive correlations with other assets.

Hong (2017) suggests that if the bitcoin returns' time-series momentum is significant to make profits, momentum trading on bitcoin will bring higher and non-correlated returns to the portfolio. Hong's study suggests high diversification effects-high bitcoin momentum returns, but the momentum cycle presents a faster pattern than other assets.

The risk of bitcoin investment appears to be challenging to handle. Firstly, bitcoin prices often vary widely with large swings, which makes the buy-and-hold strategy very risky and frustrates risk-averse investors. Secondly, because of the low correlation with other assets, hedging exposure to wide price variations tends to be ineffective (Yermack, 2013). Similarly, Bouri et al. (2017) found that bitcoin performs as a strong hedge and safe (hedging) heaven only in very few cases. Thirdly, the situation did not improve after the birth of bitcoin futures contracts introduced by the Chicago Mercantile Exchange (CME) and the Chicago Board Options Exchange (CBOE) in December 2017. Bitcoin futures cannot mitigate the pricing risk in the spot bitcoin market; conversely, hedging with them leads to increased volatility (Corbet et al., 2018). Fourthly, bitcoin risk management seems

to be complicated due to regime-switching. Li et al. (2018) found that several factors affecting bitcoin risk present an asymmetric pattern under different risk regimes.

2.4 Bitcoin price

The bitcoin price follows the law of supply and demand. Since bitcoin supply is growing at a predictable speed and the maximum number of bitcoins has been set at 21 million, bitcoin demand becomes the determining force in its price formation.

Several factors, including attractiveness to investors and various macroeconomic and financial developments, have been identified as affecting the bitcoin price (Buchholz et al., 2012; Kristoufek, 2013; van Wijk, 2013). However, Ciaian et al. (2016) found that the bitcoin price is not driven by macro-financial developments in the long run, when considering all factors. This result raises the point of bitcoin's decentralized architecture, independent of macro decisions. Moreover, the arrival of new information positively impacts the bitcoin price, and the alternation of positive and negative news has generated highly volatile price cycles (Ciaian et al., 2016).

The bitcoin price has a substantial speculative bubble component, and the intrinsic worth of a bitcoin is zero (Cheah & Fry, 2015). As shown in the literature, bitcoin's value relies on its usefulness as a unit of account, its attractiveness as a medium of exchange, it role as a store of value and a useful investment tool, or even fondness for its niche.

2.5 The Bitcoin exchange rate

To date, the literature about the bitcoin exchange rate remains limited. The only relevant paper that I have found was written by Dong and Dong (2015). They use a linear regression model and a Granger-causality model to show that the bitcoin exchange rate –they call it the

bitcoin parity rate – deviates from the corresponding FX rate persistently. This persistent arbitrage stickiness is interpreted as people's preference for a buy-and-hold strategy with bitcoins in their work.

Bitcoin is viewed by many as either a currency or an investment asset. Both roles present some striking features, while criticisms also prevail. Bitcoin is now welcomed by market participants as a unique investment asset. However, widely swinging bitcoin prices raise problems for both of its functions.

Nan and Kaizoji (2017, 2019b) propose to treat bitcoins as a medium of foreign exchange. In bitcoin markets, the U.S. dollar can be first changed into bitcoins, which can then be immediately traded for Euros. If performed very swiftly, this two-step switching approach reduces the risks of holding bitcoins to the minimum. The bitcoin exchange (BX) rate, more precisely the USD/EUR rate defined in equation (1), represents the U.S. dollar price of a Euro when the two switching steps take place simultaneously. (We will discuss later the reasons why the bitcoin exchange rate stands for a fair approximation of the market bitcoin-based foreign exchange rate.) They found that the USD/EUR BX rate and the USD/EUR FX rate were in long-run equilibrium over the period from 1 May 2014 to 21 November 2017, and that the latter appears to be an unbiased estimator of the former. In the short run, however, divergence from equilibrium occurs frequently, indicating a possible arbitrage opportunity.

Based on the USD/EUR BX, Nan and Kaizoji (2019a, 2020) proposed triangular arbitrage with the FX rate and showed that the risks of trading on the BX rate or the triangular arbitrage can be effectively managed using the FX futures contract as a hedge.

In the Ethereum markets, similar results regarding the ether-based exchange rate, long-run market equilibrium with the corresponding FX rate, and the short-run adjustment dynamics have been found by Pichl, Nan, and Kaizoji (2020).

3 Characterization

This chapter describes the data set and presents a statistical description and characterization of the time series.

3.1 Data

The data set, provided by Bloomberg, consists of two types of periods: the daily period and the 5-minute period. For most of the hypothesis tests, modeling, and forecasts, we use the daily data. The 5-minute data are used for forecasting with neural networks and discussing the time delay problem in bitcoin transactions. The daily data set covers the period from 10 September 2013 to 06 March 2020. The range of the 5-minute data set is between 2014/05/01 10:05 JST and 2020/03/08 9:55 JST. Bloomberg only provides 5-minute data beginning 1 May 2014.

For bitcoin prices, this thesis uses two bitcoin indices: the bitcoin index of USD (USD/BTC), a composite index of closing prices from four bitcoin markets (Bitstamp, Coinbase, itBit, and Kraken), and the bitcoin index of EUR (EUR/BTC) from the itBit and Kraken markets.

The exchange rates from the FX market include the USD/EUR spot rate and the one-month futures rate, quoted in closing prices. The one-month futures rate is the USD/EUR

futures contract in the monthly cycle listed on the Chicago Mercantile Exchange. The one-month forward rate is calculated from the forward points by multiplying them by 0.0001 and adding the spot rate.

The one-month deposit rates include the one-month USD ICE LIBOR interest rate and the one-month ERU ICE LIBOR interest rate. Both deposit rates are quoted bids to maintain the practical sequence of arbitrage, i.e., an arbitrager can lend either U.S. dollars or Euros at the bid sides of the markets. ICE LIBOR is a benchmark rate that represents the interest rate at which banks offer to lend funds to one another in the international interbank market for short-term loans; it is an average value of the interest rate calculated from estimates submitted by the leading global banks on a daily basis (Bloomberg).

The USD/EUR bitcoin exchange rate is calculated using Equation (1), that is, $(USD/EUR)^{BX} = (USD/BTC)/(EUR/BTC)$, which describes the process in which a market participant trades a Euro for bitcoins, and then trades the bitcoins for U.S. dollars, denoted by $EUR \rightarrow BTC \rightarrow USD$.

The triangular arbitrage between the USD/EUR BX rate and the USD/EUR FX rate, defined in (Nan & Kaizoji, 2019a), is expressed by

$$Triangular\ arbitrage = (USD/EUR)^{BX} \times EUR/USD \tag{2}$$

where EUR/USD=1/(USD/EUR) is the reciprocal of the USD/EUR spot rate. The triangular arbitrage process is denoted by $EUR \rightarrow BTC \rightarrow USD \rightarrow EUR$.

Pichl & Kaizoji (2017) define the profit rate relative to bitcoin price in Euros as $(2) - 1$. We modify their idea and define buying one dollar in euros in the bitcoin markets as $-(EUR/USD)^{BX}$ (the expense) and exchanging the received dollar for Euros in the FX market as EUR/USD (the profit). The proposed profit rate is expressed as

$$\delta_{e,u} = \frac{-(EUR/USD)^{BX} + EUR/USD}{(EUR/USD)^{BX}} = -1 + (2).$$

This profit can be approximated using the difference in logarithms, as $\log(x_2) - \log(x_1) \approx (x_2 - x_1)/x_1$. So $\delta_{e,u} \approx \log(EUR/USD) - \log((EUR/USD)^{BX}) = \log((USD/EUR)^{BX}) - \log(USD/EUR)$, equal to the logarithm of equation (2). This result shows that the logarithm of equation (2) is the return on the triangular arbitrage.

Table 1 gives the series names and their descriptions.

Table 1. Series names and their descriptions

Series names	Descriptions	Number of Obvs.	Start	End
ub	The daily bitcoin index of USD	1601	2013/09/10	2020/03/06
eb	The daily bitcoin index of EUR	1601	2013/09/10	2020/03/06
ue	The daily USD/EUR FX spot rate	1601	2013/09/10	2020/03/06
fo_ue	The daily 1-month USD/EUR FX forward rate	1601	2013/09/10	2020/03/06
fu_ue	The daily 1-month USD/EUR FX futures rate	1601	2013/09/10	2020/03/06
bx_ue	The daily USD/EUR BX rate	1601	2013/09/10	2020/03/06
li_u	The daily 1-month USD LIBOR interest rate	1601	2013/09/10	2020/03/06
li_e	The daily 1-month EUR LIBOR interest rate	1601	2013/09/10	2020/03/06
ub_5m	The 5-minute bitcoin index of USD	364937	2014/05/01 10:05 JST	2019/03/08 9:55 JST
eb_5m	The 5-minute bitcoin index of EUR	364937	2014/05/01 10:05 JST	2019/03/08 9:55 JST
ue_5m	The 5-minute USD/EUR FX spot rate	364937	2014/05/01 10:05 JST	2019/03/08 9:55 JST
bx_ue_5m	The 5-minute USD/EUR BX rate	364937	2014/05/01 10:05 JST	2019/03/08 9:55 JST

Note: The 5-minute data cover a shorter period than the daily data due to data availability.

Figure 2 gives the plots of the series: the bitcoin indices of USD and EUR, plotted respectively, and the bitcoin exchange rate of USD/EUR and the FX spot of USD/EUR, plotted jointly. As shown in panel (c) of Figure 2, the bitcoin exchange rate of USD/EUR (bx_ue) oscillated before May 2014, but gradually mimicked the FX spot, except for several spikes. Interestingly, though the behavior of the bitcoin price indices was very volatile at the end of 2018, the bitcoin exchange rate did not show a substantial divergence from the FX spot.

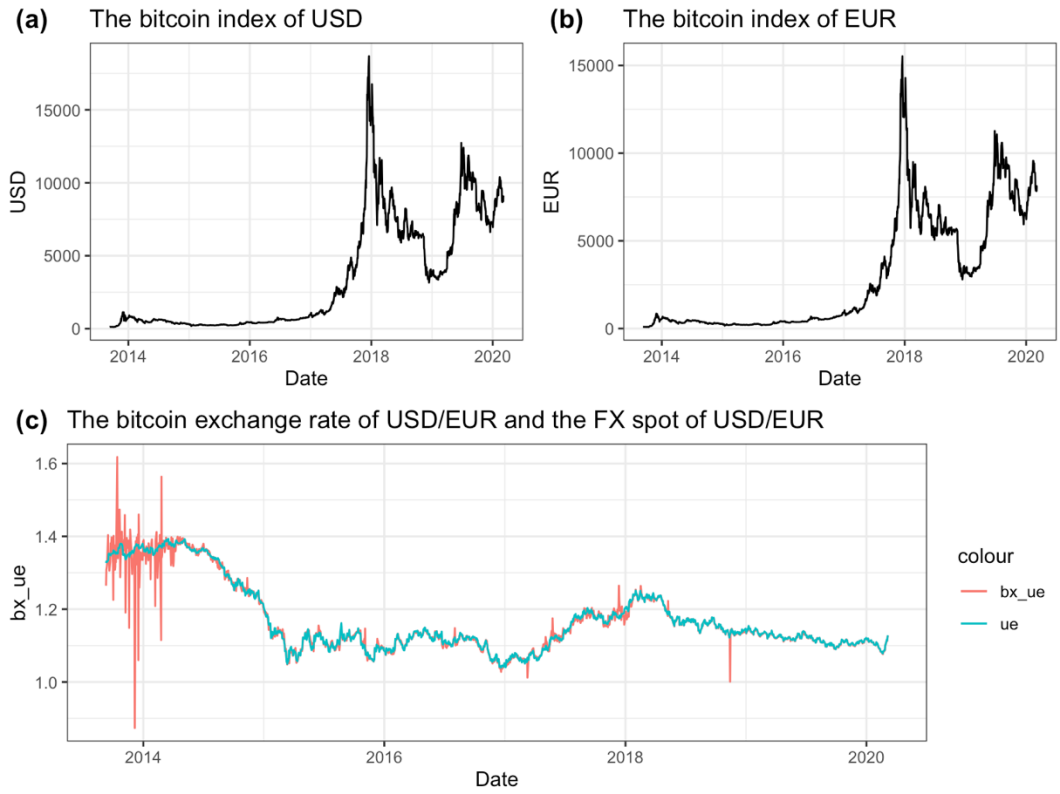

Figure 2. The bitcoin indices of USD and EUR, the bitcoin exchange rate of USD/EUR (bx_ue), and the FX spot of USD/EUR (ue) (10 September 2013 – 06 March 2020).

3.2 Statistical descriptions

This section investigates the statistical features of the time series in the data set. The summary statistics in Table 2 give a statistical description of the level of each time series over its sample period. Bitcoin prices varied within a wide range; for instance, USD/BTC (ub) fluctuated between $112.8 and $18,674.5. The highest USD/BTC price ($19,462.2) occurred in an intraday transaction, as shown in the 5-minute data. In contrast, the USD/EUR bitcoin exchange rate varied between 0.874 and 1.618, showing more volatility than the USD/EUR FX rate (1.039, 1.393). Second, the standard deviation of USD/BTC, indicating variation from the mean, reached 3,830.4, whereas, the S.D. of the USD/EUR BX rate is 0.095, close to the 0.093 S.D. of the FX spot. The median value of USD/BTC indicates that

half of the sample values are lower than $869.9. The significant difference between the mean and the median implies a strongly left-skewed distribution of the level bitcoin price of USD.

Table 2. Summary statistics of the level data

Series	Obvs.	Mean	Median	Min.	Max.	S.D.
ub	1601	3389.9	869.9	112.8	18674.5	3830.4
eb	1601	2944.5	752.0	84.0	15528.9	3323.0
ue	1601	1.170	1.134	1.039	1.393	0.094
fo_ue	1601	1.172	1.136	1.041	1.393	0.093
fu_ue	1601	1.172	1.136	1.043	1.394	0.093
bx_ue	1601	1.169	1.133	0.874	1.618	0.095
li_u	1601	1.014	0.695	0.148	2.522	0.847
li_e	1601	-0.250	-0.381	-0.573	0.249	0.225
ub_5m	364937	2741.5	674.1	160.4	19462.2	3489.6
eb_5m	364937	2333.8	610.3	148.7	15939.6	2910.7
ue_5m	364937	1.155	1.136	1.034	1.395	0.075
bx_ue_5m	364937	1.154	1.136	0.870	1.413	0.075

Note: Obvs. refers to the number of observations; S.D. refers to the standard deviation.

Note that the summary statistics of the level data are not very meaningful if the series is a random walk process, as these summary statistics are always changing as new samples arrive. However, the return of a random walk series has a mean-reverting tendency, meaning that the summary statistics on the return series will be more informative.

Moreover, stock prices are said to follow the log-normal distribution, so the logarithmic returns of price have a hypothetical normal distribution, which is of interest. Based on this statistical feature, the logarithmic return can be a good approximation of the return.

Table 3 gives the summary statistics of the logarithmic returns of the time series. Most of the mean returns are very close to zero, except for the returns of the two bitcoin indices and the triangular arbitrage. The returns of the bitcoin indices of USD and ERU both have a positive mean equal to 0.003, indicating a 0.3% risk premium for holding bitcoins.

Based on the minimum and maximum values, we find that bitcoin trading tends to face a large degree of variability relative to both FX trading and bitcoin-based currency

trading – in the extreme case, the bitcoin value may lose or gain more than 50% against the USD in a day. In contrast, the maximum daily loss and gain for USD/EUR trading in the FX market are -2.4% and 3%, respectively. The situation seems not to be alleviated through the proposed bitcoin-based currency trading: The maximum loss and gain are -37.6% and 46.2% for the BX return, and -45% and 17.6% for the triangular arbitrage return. However, the FX return has the lowest S.D., followed by the return on the triangular arbitrage and the BX rate, while the returns of the bitcoin indices have the highest value.

Table 3. Summary statistics of the logarithmic returns

Returns	**Obvs.**	**Mean**	**Min.**	**Max.**	**S.D.**	**Skewness**	**Kurtosis**
r_ub	1600	0.003	-0.584	0.517	0.054	0.026	20.938
r_eb	1600	0.003	-0.264	0.461	0.049	0.384	8.874
r_ue	1600	-0.000	-0.024	0.030	0.005	0.123	2.637
r_fo_ue	1600	-0.000	-0.024	0.030	0.005	0.120	2.630
r_fu_ue	1600	-0.000	-0.021	0.033	0.005	0.158	2.508
r_bx_ue	1600	-0.000	-0.376	0.462	0.027	2.873	123.810
r_ta	1600	-0.001	-0.450	0.176	0.019	-10.159	219.880
r_ub_5m	364936	0.000	-0.240	0.257	0.005	-0.397	313.683
r_eb_5m	364936	0.000	-0.189	0.234	0.005	0.100	328.896
r_ue_5m	364936	-0.000	-0.026	0.027	0.001	0.270	212.416
r_bx_ue_5m	364936	-0.000	-0.161	0.094	0.003	-0.870	161.906
r_ta_5m	364936	-0.000	-0.206	0.170	0.008	2.644	37.867

Note: Kurtosis refers to the excess kurtosis.
The value -0.000 indicates a negative value greater than -0.0005.
All series were tested for normality using the Jarque-Bera test; the results are suggestive of rejecting the normality null hypothesis.

All series show excess kurtosis and thus lead to rejecting the null hypothesis of normality. The results of the Jarque-Bera test also reject the null hypothesis of normality. Typically, non-normal return densities indicate weak dependence in the return series. The presence of the ARCH or GARCH in a return series may cause a leptokurtic density (Baillie & Myers, 1991).

Figure 3 plots the sample probability densities of the twelve logarithmic return series. It is apparent that the bitcoin returns show high excess kurtosis and fat tails in their probability density.

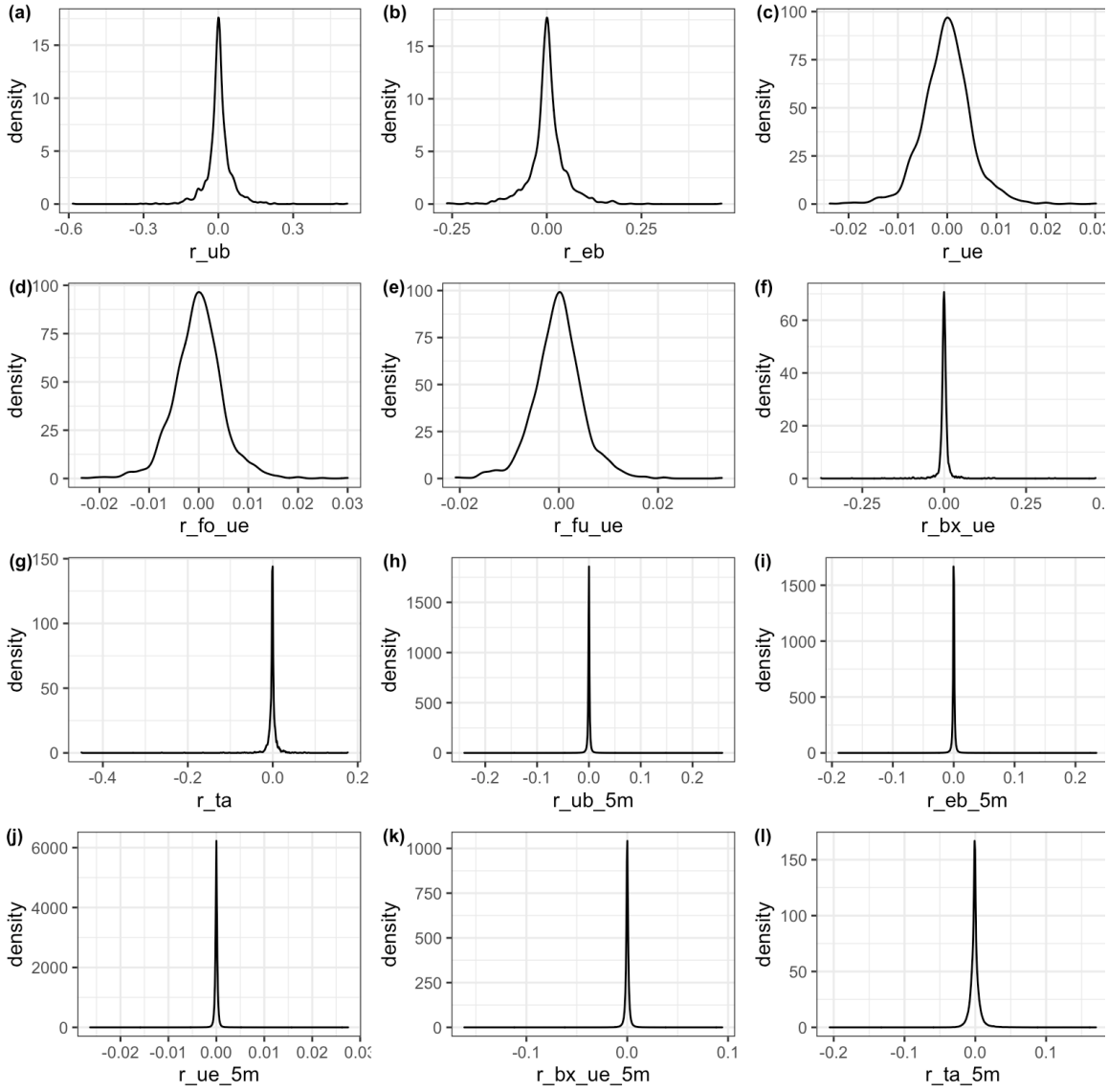

Figure 3. The sample probability densities of twelve return-series.

3.3 Time-series stylized facts

Modern time-series econometrics aims to estimate and forecast the irregular component that exists in economic data. Some of the empirical findings that were extracted from the stochastic trend are consistent across assets, markets and time. Such findings are referred to

as stylized facts. Whether these well-accepted stylized facts also adapt to bitcoin-related assets or portfolios is a question to be investigated. This section conducts time-series analyses on daily basis. Note that all data series are in natural logarithms from this section forward, though the notation remains unchanged.

3.3.1 Random walk process and the weak-form test for market efficiency.

Randomness is a concept central to financial time series. A stochastic process is expressed as a collection of random variables. One task of modern time-series econometrics is to estimate and forecast stochastic trend (Enders, 2014, p. 2). For this reason, the random walk model is the basic building block for many econometric models.

The importance of randomness also lies in its close connection with market efficiency theory. The efficient market hypothesis (EMH) posits that prices should always fully reflect all available information (Fama, 1970). Depending upon the level of available information, Fama groups market efficiency into three forms: weak, semi-strong, and strong.

Weak form tests are concerned with a subset of information containing only historical prices. Specifically, if a market is efficient in the weak form, the implications of the historical prices have been agreed upon by the market participants, so that the current price should be independent of its history. As a result, the day-to-day change in the price of a stock is entirely random.

In addition to stock prices, most macroeconomic data series are found to be a random walk process (Nelson & Plosser, 1982). The proposed bitcoin exchange rate, the bitcoin price and the triangular arbitrage call randomness into question.

Hypothesis 3.1 The USD/EUR bitcoin exchange rate series is a random walk process.

3.3.1.1 Methodology. The simplest random walk model has the form

$$y_t = y_{t-1} + \varepsilon_t \tag{3}$$

where y_t is the current logarithm of the price on day t and ε_t is a random residual term—white noise. This model suggests that (i) the daily change in price is random, that is $\Delta y_t = \varepsilon_t$, and (ii) the random residual term has a mean value of zero, that is, $E(\varepsilon_t) = 0$.

To make the random walk hypothesis testable, Dickey and Fuller (1979) consider the more general stochastic difference equation

$$\Delta y_t = \alpha_0 + a_1 y_{t-1} + \varepsilon_t \tag{4}$$

and specify the restriction $\alpha_0 = a_1 = 0$ for the random walk hypothesis. The restriction $\alpha_0 = 0$ assures that ε_t has a zero mean; $a_1 = 0$ asserts that Δy_t exhibits randomness, as ε_t and Δy_t has no correlation with the historical price y_{t-1}. Note that $\alpha_0 = 0$ is equivalent to the statement that the stochastic difference equation in (4) has a unit root. For this reason, this type of test is often called the unit root test.

This thesis employs three types of unit-root tests—the Augmented Dickey-Fuller (ADF) test, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and the Elliott-Rothenberg-Stock (ERS) test—to make inferences regarding the random walk hypothesis and its implication of weak-form EMH.

Note that evidence that the disturbance term ε_t in (4) is predictable invalidates the random walk hypothesis (Enders, 2014, p. 4). To resolve the problem that serial correlation may exist in ε_t, the ADF test modifies the DF test by adding p lags of Δy_t and a time trend:

$$\Delta y_t = \alpha_0 + a_1 y_{t-1} + a_2 t + \sum_{i=1}^{p} \beta_i \, \Delta y_{t-i} + \varepsilon_t. \tag{5}$$

where lag length p is determined using the Akaike Information Criterion (AIC).

Because the ADF test works poorly in discriminating a near-unit-root process from a unit-root process, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Elliott-Rothenberg-Stock (ERS) tests are conducted as supplementary methods.

Different from the ADF test, the KPSS test posits the null hypothesis of stationarity. It modifies the Lagrange Multiplier (LM) to first eliminate the deterministic level (de-meaned model) and the deterministic trend (de-trended model), and then tests for stationarity as the null hypothesis. The null hypothesis proposes a model composed of a deterministic linear trend, a random walk, and a stationary error term and assumes that the random walk has zero variance (Kwiatkowski et al., 1992). The lag selection here uses a "short"-type lag, $lags = \sqrt[4]{4 \times (T/100)}$, except for the BX rate series that uses a "long-type lag, $lags = \sqrt[4]{12 \times (T/100)}$.

Elliott et al. (1996) argue that the LM model in the first difference used by the KPSS test is misspecified and proposes a trend stationary (TS) model to de-mean and de-trend the series. The de-trended and de-meaned series is then used to approximate the ADF test. The ERS test is often called the Dickey-Fuller generalized least squares (DF-GLS) test.

3.3.1.2 Results. The results of the random walk tests are presented in Table 4. The ADF tests use the trend model shown in equation (5). For the trend model, the $\hat{\tau}_\tau$ statistic is used to determine whether $a_1 = 0$. The results show no significance for any of the $\hat{\tau}_\tau$ statistics, which means that we cannot reject $a_1 = 0$, the presence of a unit root. The t-statistics for the trend coefficient $a_0 = 0$ present insignificance, except in the case of the BX rate, where its value is very close to zero. The changes in the BX rate and the three FX rates all present a zero mean as indicated by the value of a_0. Note that the changes in bitcoin prices of USD and EUR show positive non-zero means, $a_0 = 2\%$. This result is similar to their mean logarithmic returns as shown in Table 3. Be cautious that the unconditional mean

of the changes in bitcoin price is not equal to a_0. For example, Δy_t of USD/BTC is an autoregressive process of order thirteen, denoted by AR (13), so its unconditional mean is given by $\frac{a_0}{B(L)}$, where L is the lag operator and $B(L)$ is a polynomial of order p of β_i in (5).

The KPSS test results show that all the $\hat{\eta}_\tau$ statistics led to the rejection of the null hypothesis, indicating acceptance of the alternative unit root hypothesis.

The ERS test applies the ADF test without an intercept to the de-meaned data. The statistics for the de-meaned model all suggest the existence of a unit root. The lag length follows the length used in the ADF test.

Table 4. Results of random walk tests (unit root tests)

Series	**ADF**				**KPSS**		**ERS**	
	Lags	α_0	α_2	$\hat{\tau}_t$	Lags	$\hat{\eta}_\mu$	Lags	$\hat{\tau}_\mu$
ub	13	0.02**	0.00	-1.97	24	5.56***	13	0.72
eb	5	0.02**	0.00	-1.94	8	15.71***	5	0.93
ue	1	0.00	-0.00	-1.59	8	6.32***	1	-0.32
fo_ue	1	0.00	-0.00	-1.59	8	6.19***	1	-0.24
fu_ue	1	0.00	-0.00	-1.61	8	6.11***	1	-0.27
bx_ue	17	0.00	-0.00*	-1.71	24	2.29***	17	-0.57

Note: $\hat{\tau}_t$ statistic of the ADF test is used to determine whether $\alpha_1 = 0$ in equation (5) that includes a time trend term. Since most of the trend-term coefficients are not statistically significant, the KPSS and ERS tests use de-meaned models. The unit root null hypothesis is used in the ADF and ERS tests, while the stationarity null hypothesis is used in the KPSS test.
*, **, and *** significant at 10%, 5%, and 1%, respectively.

The Q-statistic of the Ljung-Box test is applied to the error term of each ADF model to diagnose the model's adequacy. The null hypothesis asserts independence in the given time series. The notation Q (20) denotes a Q-statistic using a lag of 20. In the tests, all Q (20) statistics indicated no serial correlation in the error terms for the ADF models.

In conclusion, the USD/EUR BX rate and the bitcoin price indices of USD and EUR behave like a random walk process, just as the FX rates do, which suggests the weak-form market efficiency. There are two features related to this: (i) the current changes in the price of a series are uncorrelated with its price history, so knowing about the history does not help

estimate the current price since prices evolve randomly; (ii) the changes in prices have a zero mean. Though for the two bitcoin price indices, the price changes present non-zero means, the results seem not to invalidate the random walk hypothesis and the implied market efficiency, as the constant in (5) can be thought of as the risk premium for holding bitcoins.

3.3.2 Serial dependence. As suggested by market efficient theory, changes in the price of an asset (or the logarithmic returns) are expected to be random, with a zero mean. This randomness should result in serial independence; however, many of the return series show serial dependence.

For a process with a unit root, its first difference is stationary. Such a sequence is integrated of order one, denoted by $I(1)$. The merit of a stationary series is that after observing a set of its realizations, we can reasonably approximate the mean, variance, and autocorrelations using averages over a sufficiently long period of time (Enders, 2014, p. 52).

Hypothesis 3.2 The logarithmic return series of the USD/EUR BX rate is stationary.

3.3.2.1 Methodology. Similar to the random walk tests, ADF, KPSS, and ERS tests are conducted to check the logarithmic return series for stationarity. An autoregressive moving average (ARMA) model is then used for modeling each return series. The ARMA (p, q) model is specified as

$$y_t = \alpha_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \qquad (6)$$

where y_t is a stationary process.

The logarithmic returns of the series are calculated by taking first-order differences of the level series; they are then checked for stationarity. For the ADF and ERS tests, rejecting the null hypothesis indicates stationarity, whereas the KPSS test uses stationarity as the null hypothesis. Note that the triangular arbitrage between the BX rate and the FX rate (the BX-FX triangular arbitrage) is viewed as a return series—specifically, the excess return. Different from the calculation in (2), the logarithmic excess return is given by $\log(USD/EUR)^{BX} - \log(USD/EUR)$.

3.3.2.2 Results. The results presented in Table 5 suggest that all return series are stationary. For the return of the bitcoin exchange rate, the ERS test does not agree with the ADF and KPSS tests, suggesting non-rejection of the null hypothesis of a unit root. This disagreement may be caused by significant autocorrelation that exists in longer lags of the BX rate return series. All bitcoin-related returns present a longer pattern of autocorrelation than the FX returns.

Table 5. Results of stationarity (unit root tests)

Series	**ADF**			**KPSS**		**ERS**	
	Lags	α_0	$\hat{\tau}_\mu$	Lags	$\hat{\eta}_\mu$	Lags	$\hat{\tau}_\mu$
r_ub	12	0.00	-9.77***	24	0.09	12	-4.78***
r_eb	4	0.00	-15.26***	8	0.10	4	-14.83***
r_ue	1	-0.00	-28.80***	8	0.17	1	-18.99***
r_fo_ue	1	-0.00	-28.79***	8	0.17	1	-19.00***
r_fu_ue	1	-0.00	-28.14***	8	0.16	1	-18.83***
r_bx_ue	16	-0.00	-12.24***	24	0.06	16	-0.64
r_ta	11	-0.00*	-8.12***	24	0.30	11	-3.47***

Note: $\hat{\tau}_\mu$ statistic of the ADF test is used to determine whether $\alpha_1 = 0$ with a draft term in the model. The KPSS and ERS tests use de-meaned models. The ADF test and the ERS test have the null hypothesis of a unit root while the KPSS test have the null hypothesis of stationarity.
*, **, and *** significant at 10%, 5%, and 1%, respectively

The order of ARMA (p, q) is first suggested by the AIC criterion, which chooses the model that has the smallest sum of squared residuals (SSR). The results are shown in Table 6.

For the return of the bitcoin price of USD, r_ub, the AIC suggests the ARMA (2, 2) model. Because a_1 and β_1 of the ARMA (2, 2) model appear to be statistically insignificant, we pare down the model by eliminating the first AR term and the first MA term. However, the log-likelihood (LL), AIC and BIC values reported in the third column of Table 6 do not show any improvement with this specification. The Q (20) statistics suggest that the ARMA models of r_ub are problematic, with significant residual autocorrelation left in 20 lags. For r_eb, ARMA (1, 2) appears to be adequate for capturing time dependence.

Table 6. Estimates of the ARMA (p, q) model of the return series.

	r_ub		**r_eb**	**r_ue**	**r_bx_ue**		**r_ta**	
	ARMA (2, 2)	ARMA [2; 2]	ARMA (1, 2)	ARMA (0, 0)	ARMA (2, 3)	ARMA [2; 1,2,3]	ARMA (5, 2)	ARMA [2, 3, 5; 2]
a_0	0.00** (0.00)	0.00* (0.00)	0.00* (0.00)				-0.00** (0.00)	-0.00** (0.00)
a_1	-0.10 (0.07)		0.87*** (0.06)		-0.33 (0.22)		0.04 (0.11)	
a_2	-0.87*** (0.04)	0.83*** (0.06)			0.31* (0.19)	0.55*** (0.14)	0.53*** (0.10)	0.57*** (0.10)
a_3							0.06** (0.03)	0.09*** (0.02)
a_4							0.04 (0.03)	
a_5							0.06** (0.03)	0.07** (0.03)
β_1	0.04 (0.08)		-0.89*** (0.06)		-0.44** (0.21)	-0.74*** (0.02)	0.01 (0.11)	
β_2	0.85*** (0.05)	-0.78*** (0.07)	0.07*** (0.03)		-0.64*** (0.09)	-0.64*** (0.14)	-0.56*** (0.10)	-0.57*** (0.10)
β_3					0.33** (0.14)	0.50*** (0.10)		
LL	2425	2417	2584	6195	3915	3913	4071	4067
AIC	-4838	-4827	-5158	-12388	-7818	-7817	-8123	-8122
SBC	-4806	-4806	-5131	-12383	-7786	-7790	-8047	-8090
Q (20)	47***	65***	7.13	24	193***	191***	194***	215***

Note: The order of the ARMA (p, q) model is suggested by the AIC criterion. Then, the ARMA model is pared down if there exist insignificant estimates. The notation [;] is used for specifying the lags in the ARMA model; inside the brackets, the AR terms and the MA terms are separated by a semicolon.
LL denotes the log-likelihood; SBC (or BIC) denotes the Schwarz's Bayesian criterion. Q (20) refers to the Q-statistic of the Ljung-Box test using a lag of 20.
*, **, and *** significant at 10%, 5%, and 1%, respectively. Standard errors are given in parentheses.

As presented in column 5 of Table 6, the FX spot return does not show significant autocorrelation in its residuals. The AIC suggests that there is no need to use an ARMA

model on the series. This result indicates an efficient FX spot market, meaning that its prices do not provide much useful forecasting information.

The return series of the BX rate is modeled by the ARMA (2, 3) using the suggestion of the AIC. After removing the first AR term, the reduced model is selected by the BIC criterion, which prefers parsimony. For the return of the triangular arbitrage, the AIC selects the ARMA (5, 2) model. The modified model remaining lag 2, 3, 5 in the AR terms, and lag 2 in the MA terms is preferred by the BIC. However, the Q (20) statistics indicate that the residuals from these models exhibit substantial serial autocorrelation.

The values of a_0 suggest that only the BX return and the FX return have a zero mean. The returns of two bitcoin prices have tiny positive means. The triangular arbitrage return shows a small negative mean.

In short, the bitcoin-related asset returns all exhibit long memory, while the FX return does not. This long-range correlation results in the high order of the ARMA model. Even so, the resulting ARMA models appeared to be inadequate, as suggested by the Q (20) statistics of the Ljung-Box tests.

3.3.3 Volatility. Many series of financial returns exhibit time-dependent variances (conditional heteroskedasticity), i.e. periods of volatility clustering followed by periods of relative calm. Just like modeling a conditional mean, an AR model is suggested to measure the time dependence of volatility. Engle (1982) proposes a model, called the autoregressive conditional heteroskedastic (ARCH) model, that can measure the time-varying mean and variance simultaneously. Bollerslev (1986) extends Engle's work by incorporating moving average terms, calling it the Generalized ARCH (GARCH) model.

The conditional heteroskedasticity will cause the probability density of the return to exhibit excess kurtosis and fat tails (Baillie & Myers, 1991). We have found that the bitcoin-

related returns all present substantially high leptokurtosis. Baillie & Myers (1991) argue that the hypothetical Student's t distribution works better than the normal distribution for GARCH models in the presence of excess kurtosis. Nelson (1991) proposed the generalized error distribution (GED) with parameter ν measuring tail-thickness. When $\nu = 2$, the GED converges to the standard normal distribution; for $\nu < 2$, the distribution has heavier tails than the normal, while for $\nu > 2$, it has thinner tails, finally converging to the uniform distribution (Angelidis et al., 2004).

Hypothesis 3.3 The return series of the BX rate is conditional heteroskedastic.

3.3.3.1 Methodology. A GARCH (p, q) model is given by

$$\varepsilon_t = v_t\sqrt{h_t}$$
$$h_t = \omega_0 + \sum_{i=1}^{q} \omega_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \theta_i h_{t-i} \tag{7}$$

where ε_t is the residual process from the ARMA model specified in (6), v_t is a white-noise process with variance of 1.0, and h_t is the conditional variance of the return (Enders, 2014, p. 128).

With the results from Section 3.3.2, an ARMA model and a GARCH (1, 1) model are jointly estimated using the maximum likelihood estimation (MLE). Before the estimation, it is recommended that the Lagrange multiplier (ML) test be used to diagnose ARCH or GARCH effects. The test statistic TR^2 converges to a χ^2 distribution with q degrees, where q is the number of lags in the regression (Enders, 2014, p. 130). Only the first 1000 observations are used for the estimation; the remaining 600 values are retained for forecasting.

For comparison, the proposed models are estimated using three types of densities: the normal distribution, the Student's t distribution and the GED. The ARMA (1, 1) plus GARCH (1, 1) model is also estimated as a trial.

For comparison, the proposed models are estimated using three types of densities: the normal distribution, the Student's t distribution and the GED. The ARMA (1, 1) plus GARCH (1, 1) model is also estimated as a trial.

Table 7. The estimation of the ARMA (p, q) + GARCH (1, 1) model: the USD/EUR BX rate.

	Model 1 ARMA (1, 1) GARCH (1, 1) Norm.	Model 2 ARMA (2, 3) GARCH (1, 1) Norm.	Model 3 ARMA (2, 3) GARCH (1, 1) Norm.	Model 4 ARMA [2; 1, 2, 3] GARCH (1, 1) Stud. t	Model 5 ARMA [2; 1, 2, 3] GARCH (1, 1) GED
a_0	-0.00 (0.00)	-0.00 (0.00)			
a_1	0.03 (0.16)	0.43*** (0.05)	-0.84*** (0.05)		
a_2		0.53*** (0.04)	-0.95*** (0.07)	0.92*** (0.07)	0.90*** (0.02)
β_1	-0.195 (0.138)	-0.59*** (0.00)	0.70*** (0.06)	-0.22*** (0.04)	-0.19*** (0.01)
β_2		-0.51*** (0.01)	0.82*** (0.15)	-0.93*** (0.05)	-0.92*** (0.01)
β_3		0.13*** (0.03)	-0.11*** (0.04)	0.21*** (0.04)	0.19*** (0.00)
ω_0	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ω_1	0.196*** (0.07)	0.199*** (0.07)	0.232*** (0.08)	0.446*** (0.12)	0.380*** (0.16)
θ_1	0.803*** (0.07)	0.080*** (0.07)	0.767*** (0.16)	0.551*** (0.12)	0.615*** (0.12)
ν_{std}				3.39*** (0.49)	
ν_{ged}					0.97*** (0.07)
LL	3120.62	3122.36	3127.89	3248.18	3234.14
AIC	-6.23	-6.23	-6.24	-6.48	-6.45
BIC	-6.20	-6.18	-6.20	-6.44	-6.41
$Q_{sr}(\,)$	(1) 7.78*** (5) 10.24*** (9) 12.20***	(1) 8.54*** (14) 13.11*** (24) 18.26***	(1) 9.30*** (14) 15.40*** (24) 20.95***	(1) 0.86 (14) 6.75 (24) 13.90	(1) 2.28 (14) 7.65 (24) 14.09
$Q_{ssr}(\,)$	(1) 8.41*** (5) 9.78*** (9) 10.82**	(1) 8.53*** (14) 9.93*** (24) 11.00**	(1) 7.46*** (14) 9.02** (24) 10.16**	(1) 4.43 (14) 5.28 (24) 6.29	(1) 2.99 (14) 4.01 (24) 4.73
LM (3)	0.80	0.79	0.88	0.54	0.61
LM (5)	2.42	2.48	2.68	1.39	1.58
LM (7)	2.83	2.90	3.11	2.10	1.90

Note: The ARMA (p, q) and GARCH (1, 1) model are specified in (6) and (7), respectively; The [;] notation specifies the lags in AR terms and MA terms, separated by the semicolon.

Norm. stands for the normal distribution and Stud. t stands for student's t distribution. Parameter ν_{std} is the shape parameter of the Stud. t and ν_{ged} is the shape parameter for the GED.
LL denotes the log-likelihood; $Q_{sr}(\,)$ is the Q-statistic of the Ljung-Box test on standardized residuals; $Q_{ssr}(\,)$ is the Q-statistic on square of standardized residuals; LM ( ) is the Lagrange multiplier test; inside the parentheses is the number of lags.
*, **, and *** significant at 10%, 5%, and 1%, respectively. Robust standard errors are given in parentheses below the estimated values.

3.3.3.2 Results. Table 7 shows the estimated ARMA (p, q) plus GARCH (1, 1) models on the USE/EUR bitcoin exchange rate. Model 4 in column 5 of Table 7 dominates the other models. Its ARMA part includes AR2 and MA1-3 terms, and it uses the Student's t as the hypothetical distribution.

The model selection process is illustrated as follows: Starting from the ARMA (1, 1) model under the normal distribution is to verify whether the MLE method will produce the same result as the OLS regression. As presented in column 2 of the table, none of the coefficients in the ARMA (1, 1) are significantly different from zero. The t-statistics of the GARCH (1, 1) part are all statistically significant at the 1% level and the sum $\omega_1 + \theta_1 = 0.196 + 0.803 = 0.999 < 1$ indicates that the conditional variance of ε_t is convergent but very persistent. From the values of ω_1 and θ_1, we learn that a one-unit increment of ε_t^2 increases the current conditional variance h_t by 0.196, while an increment of h_{t-1} raises h_t by 0.803. The adequacy checks suggest that serial autocorrelation exists both in the level residuals and the squared residuals, since the Q_{sr} and Q_{ssr} statistics are all significant at the 1% level. Note that the Q_{ssr} statistic tests for quadratic serial autocorrelation in the residuals indicating nonlinearity. The TR^2 statistics of the LM tests imply that there are no ARCH effects left in the residuals.

In comparison, Model 2, in the third column of the table, uses ARMA (2, 3) for capturing the conditional mean. The value of α_0 appears to be insignificant. The Q_{sr} and Q_{ssr} statistics suggest Model 2 is also not adequate; however, there are no ARCH effects as

indicated by the LM tests. Nevertheless, Model 2 seems not competitive with Model 1 because only the log-likelihood agrees with Model 2; both the AIC and BIC suggest no improvement on Model 2. These results suggest that the augmented ARMA (2, 3) model increases the log-likelihood but does not improve the SSR of the model, while at the same time increasing some costs due to the addition of more parameters.

Model 3 in column 4 of the table removes the intercept in its ARMA (2, 3) specification. The log-likelihood, AIC, and BIC all coincide in this modification, presenting an improvement. This result shows that both the MLE and OSL approaches converge to the same selection of the model. However, adequacy occurs in all three models with a normal distribution.

Model 4 uses the ARMA [2; 1, 2, 3] terms; that is, it omits the intercept and the AR1 term, and incorporates a Student's t distribution to adapt to the heavy tails. This model tends to be superior to the other models, as the log-likelihood, AIC, BIC all present a significant improvement; serial autocorrelation disappears from both the standardized residuals and the squares of standardized residuals. The values of ω_1 and θ_1 indicate that a one-unit change in ε_t^2 contributes 0.446 to the change in h_t , while a change in h_{t-1} contributes 0.551. The sum $\omega_1 + \theta_1 = 0.446 + 0.551 = 0.997 < 1$ indicates convergence and less persistence than in Model 1.

Model 5 in column 6 of Table 7 using GED as the hypothetical distribution appears to be a good approximation but is no better than Model 4, as suggested by the Log-likelihood, AIC, and BIC.

Figure 4 plots the estimated conditional standard deviation series from Model 4 over the sample period, together with the unconditional standard deviations and the 150-day moving average. Since volatility is latent, the absolute returns of the BX rate are used as a

proxy, shown as the grey line in Figure 4. The BX return was extremely volatile before May 2014, affecting the unconditional standard deviation and shifting it upward over the entire estimating period. This issue caused an overestimated risk level afterwards. As for the moving average of the standard deviation, it was sluggish and remained at a very low level for most of the time, seemingly unaffected by the frequent spikes. The conditional standard deviation appears to be most realistic, varying with the absolute returns.

For the triangular arbitrage series, consider the ARMA [2; 1, 2] plus GARCH (1, 1) model with the Student's t distribution. Though the ARMA specification is different from the ARMA (5, 2) model that we have estimated in Table 6, this model provides a better fit. The estimated model is given by

$$\begin{aligned} r\_ta &= \underset{(0.00)}{-0.001} + \underset{(0.05)}{0.657} r\_ta_{t-2} + \varepsilon_t + \underset{(0.03)}{0.385}\varepsilon_{t-1} - \underset{(0.05)}{0.391}\varepsilon_{t-2} \\ h_t &= \underset{(0.00)}{0.000} + \underset{(0.12)}{0.350}\varepsilon_{t-1}^2 + \underset{(0.06)}{0.649} h_{t-1}. \end{aligned} \tag{8}$$

The shape parameter is equal to 2.91, significant at the 1% level, indicating heavier tails. The Ljung-Box tests show no presence of serial autocorrelation both in the standardized residuals and the squares of the standardized residuals. The LM tests indicate that no ARCH effects are left in the residuals.

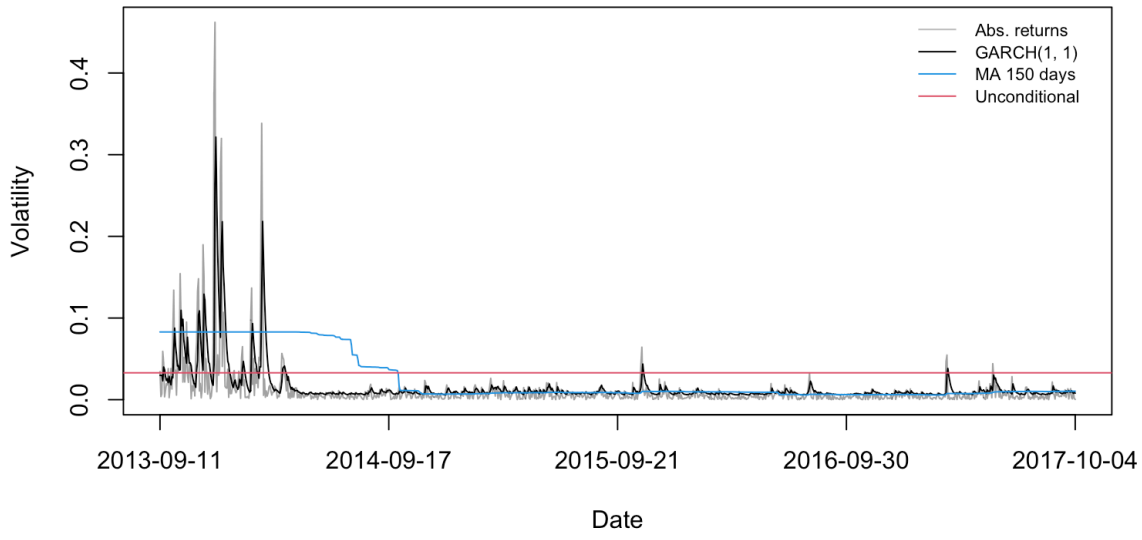

Figure 4. The estimations of volatility on the USD/EUR BX rate: the absolute returns in the grey line, the conditional standard deviation (S.D.) of ARMA [2; 1, 2, 3] plus GARCH (1, 1) in the black line, the 150-day moving average of S.D. in the blue line, and the unconditional S.D. in the red line.

The estimation on the FX spot suggests that the GARCH (1, 1) model with the Student's t distribution is adequate. The model is specified as

$$h_t = \underset{(0.00)}{0.000} + \underset{(0.01)}{0.042} r\_ue_{t-1}^2 + \underset{(0.01)}{0.956} h_{t-1}. \tag{9}$$

The shape parameter is equal to 6.54, significant at the 1% level, indicating leptokurtosis but less than in the case of the triangular arbitrage series.

For the returns of the bitcoin price index of USD/BTC, the GARCH (1, 1) with the GED provides the best fit according to the log-likelihood, AIC, and BIC. There are no GARCH effects left as indicated by the LM tests and the Ljung-Box test on the squared residuals, but serial correlation exists in the level residuals. The shape parameter of GED is equal to 0.80, which is significant at 1%, and indicates thicker tails. The model is given by

$$r\_ub = \underset{(0.00)}{0.002} + \varepsilon_t$$
$$h_t = \underset{(0.00)}{0.000} + \underset{(0.05)}{0.202}\varepsilon_{t-1}^2 + \underset{(0.06)}{0.797}h_{t-1}. \quad (10)$$

3.3.4 Value at Risk. Value-at-Risk (VaR) is a concept used by portfolio managers to measure the downside outcome of a portfolio at a given confidence level over a period. VaR estimates the tails of the empirical distribution. There are several different approaches to calculate VaR from the data. For example, if a daily return of a portfolio r_t follows a normal distribution with mean μ and standard deviation σ, denoted by $r_t \sim N(\mu, \sigma)$, VaR for one day at the α% confidence level is defined as

$$VaR(\alpha) = \mu + \sigma N^{-1}(\alpha) \quad (11)$$

where $N^{-1}(\cdot)$ denotes the inverse of the cumulative normal density function. Note that the position of portfolio is assumed to be unity.

Hypothesis 3.4 The bitcoin exchange rate presents lower risks than the bitcoin prices.

3.3.4.1 Static VaR. We first consider two static approaches over the full sample period (1600 observations)—the historical VaR and the mean-modified VaR (Favre & Galeano, 2002), which takes skewness and kurtosis into account through the use of a Cornish Fisher expansion.

Table 8. The static VaR values in the sample period.

	S.D.	Historical VaR (2.5%)	Historical VaR (1%)	Mean-modified VaR (2.5%)	Mean-modified VaR (1%)
r_ub	0.054	-0.105	-0.137	-0.179	-0.384
r_eb	0.048	-0.101	-0.131	-0.112	-0.194
r_ue	0.005	-0.010	-0.014	-0.011	-0.014
r_fo_ue	0.005	-0.010	-0.014	-0.011	-0.014

r_fu_ue	0.005	-0.010	-0.014	-0.010	-0.014
r_bx_ue	0.027	-0.027	-0.052	-0.211	-0.695
r_ta	0.019	-0.020	-0.043	-0.133	-0.434

Note: S.D. is the unconditional standard deviation. Mean-modified VaR takes skewness and kurtosis into account using a Cornish Fisher expansion.

The results are reported in Table 8. The historical VaR at the 2.5% probability level is the 2.5%-quantile of the negative returns, which is listed in column 2 of Table 8. We can find that one bitcoin will fall either in USD value or EUR value by more than 10% over a one-day period, followed by the return of the BX rate with a -2.7% one-day loss, the return of the triangular arbitrage with a -1.9% loss, and the returns of the three FX rates with a -1% loss. The historical VaR at 1% presents similar results in that the returns of the BX rate and triangular arbitrage are lower than that of bitcoin prices but higher than that of the FX rates. In contrast, the mean-modified VaR increases the downside risks of the BX rate and the triangular arbitrage when a non-normal distribution of the returns is considered. As shown in column 6 of the table, with a 1% chance, the potential loss of the BX rate is 69.5% if the hypothetical non-normal distribution is correct.

3.3.4.2 VaR with GARCH (1, 1). One problem with VaR calculated using static approaches is that it ignores volatility clustering, so that the VaR limits are breached across time. As a result, risk is underestimated during a crisis. A solution to this problem is to use the conditional standard deviation from the GARCH (1, 1) model. The VaR with GARCH (1, 1) model is given by

$$VaR(\alpha) = \mu + h^{1/2} STD^{-1}(\alpha;\, \nu)$$
$$h_t = \omega_0 + \omega_1 \varepsilon_{t-1}^2 + \theta h_{t-1} \tag{12}$$

where $STD^{-1}(\cdot)$ denotes the inverse of the Student's t density function with shape parameter ν.

Using the bitcoin exchange rage as an example, we illustrate the 2.5% VAR models based on three different standard deviations: the unconditional S.D., the 150-day moving average, and the conditional S.D. from the ARMA [2; 1, 2, 3] + GARCH [1, 1] model (see Figure 5). As can be seen from the plot, the VaR with GARCH (1, 1) (black line) is varying and lowers the VaR limit when volatility clustering occurs, whereas the limit of the static VaR (red line) and the moving average VaR (blue line) are breached.

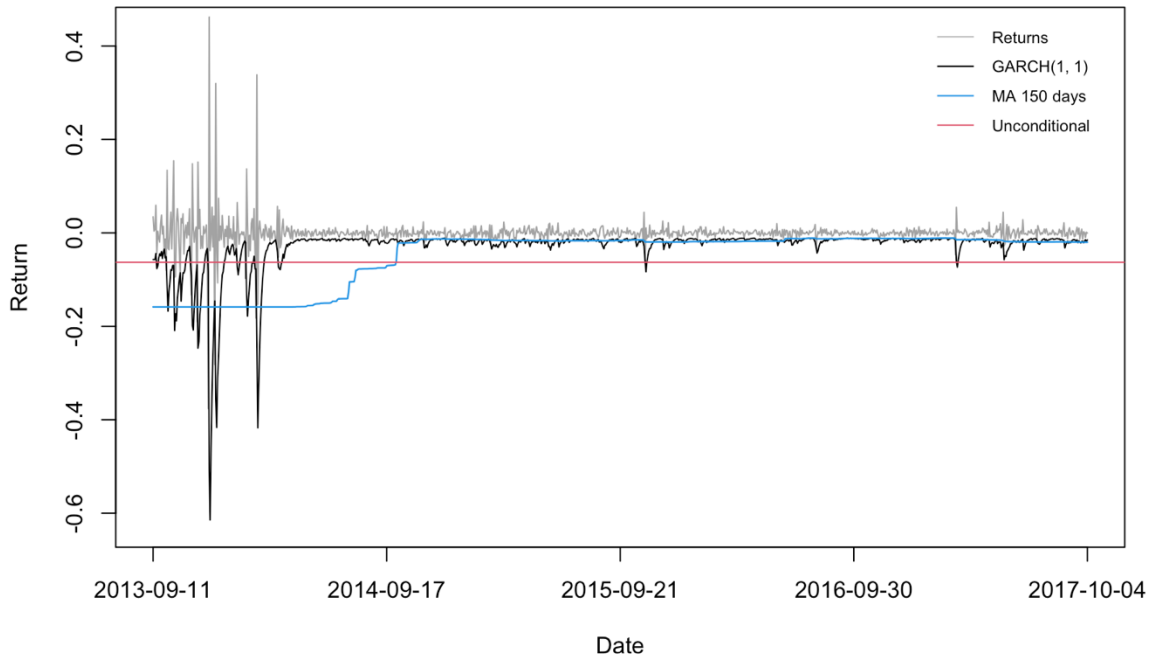

Figure 5. The return series of the USD/EUR BX rate and the three types of 2.5% VaR based on Student's distribution over the first 1000-observation sample period: the VaR with GARCH (1, 1), the 150-day moving average VaR, and the VaR with the unconditional standard deviation.

The return of the BX rate swung widely before April 2014 and then became less volatile. As a result, the static VaR becomes unrealistic when significant volatility clustering occurs, as it either overestimates or underestimates the risks. For example, the mean modified VaR suggests that by a 2.5% chance the daily loss for the BX rate is approximately 21.1%. From Figure 5, we can see that, after April 2014, the probability of such a loss

occurring appears to be zero. It seems reasonable to consider that a structural change occurred and split the sample period into two or more regimes.

3.3.5 Structural change and regime switching. The USD/EUR BX rate manifests a structural change. If we know the date when the break occurs, for example, on 1 April 2014, it is straightforward to conduct a Chow test. Once the date is confirmed by the test, the sample period can be split into two regimes: volatile and tranquil.

It is more interesting to investigate breaks that exist in the triangular arbitrage series, i.e., the excess value, defined as the difference between the log of the BX rate and the log of the FX spot. The law of one price states that in frictionless markets, the price of an identical asset will be the same, regardless of location. The USD/EUR BX rate, though it is constructed using bitcoins as the numeraire, is expected to follow the law of one price, equal to the FX spot. Any persistent deviations from the FX spot will lead to a risk-free return. Since both the bitcoin and FX markets are thought of as frictionless, the arbitrage opportunity will eventually eliminate the discrepancy between the BX rate and the FX spot. The excess value series depicts this discrepancy (see Figure 6). The breaks existing in the excess return series help us split the BX rate series into different regimes. As we know that the FX spot series presents volatility clustering, the separation based on the excess value will remove this kind of influence of the FX market so that we can focus on the transformation of the bitcoin market.

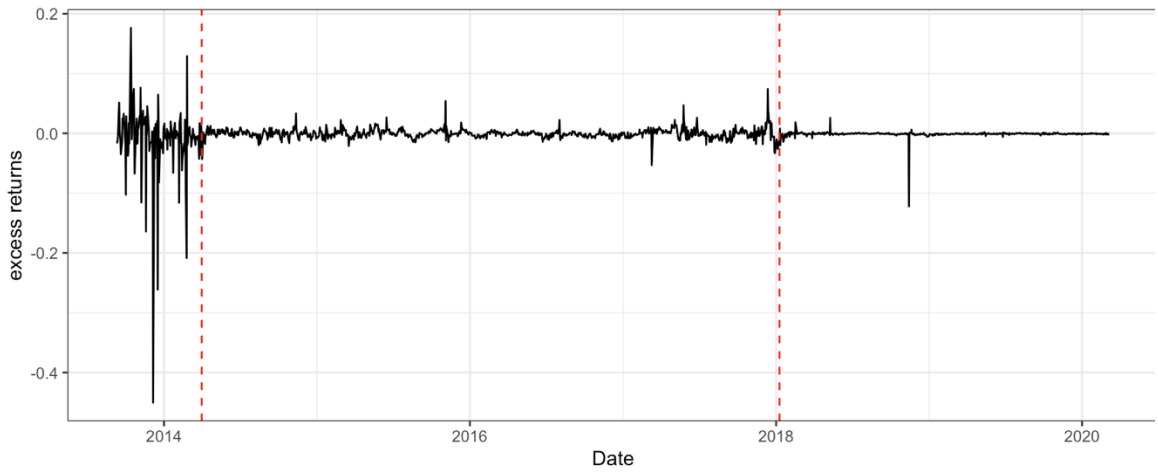

Figure 6. The logarithmic excess return for the pair of the USD/EUR BX rate and the FX spot, with two potential endogenous breaks indicated by the red lines.

Hypothesis 3.5 The returns of the triangular arbitrage (the excess values) have endogenous structural breaks.

3.3.5.1 Methodology. An endogenous break refers to a break in the time series occurring at a date that is not predetermined by the researcher (Enders, 2014, p. 104). It is natural to extend the idea of the Chow test to calculate the F-statistic for every potential bread date. The series of the excess value, r_ta, has been estimated with the ARMA (5, 2) model with a sample size of $T = 1600$ observations. If we suppose a potential break date t_m, we can use t_m to split the sample into two subsamples, with t_m observations in the first and $t_n = T - t_m$ observations in the second. Enders suggests estimating the two models

$$\begin{gathered} r_{ta_t} = \alpha_0(1) + a_1(1) r_{ta_{t-1}} + \cdots + a_5(1) r_{ta_{t-5}} + \varepsilon_t + \beta_1(1)\varepsilon_{t-1} + \beta_2(1)\varepsilon_{t-2} \\ \text{using } t_1, \ldots, t_m \\ r_{ta_t} = \alpha_0(2) + a_1(2) r_{ta_{t-1}} + \cdots + a_5(2) r_{ta_{t-5}} + \varepsilon_t + \beta_1(2)\varepsilon_{t-1} + \beta_2(2)\varepsilon_{t-2} \\ \text{using } t_{m+1}, \ldots, t_T. \end{gathered} \tag{13}$$

Then, we can use the F-test to check the restriction that all coefficients in (13) are equal. If we denote the sum of the squared residuals as SSR, the F-statistic for the restriction is given by

$$F = \frac{(SSR - SSR_1 - SSR_2)/n}{(SSR_1 + SSR_2)/(T - 2n)} \tag{14}$$

where $n = p + q + 1$ if the intercept is incorporated and the number of degrees of freedom (df.) are $(n, T - 2n)$.

Alternatively, the structural-break model can be specified using the dummy variable D_t:

$$r\_ta_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, r\_ta_{t-i} + \left(\gamma_0 + \sum_{i=1}^{p} \gamma_i r\_ta_{t-i} \right) D_t + \varepsilon_t \tag{15}$$

where dummy variable D_t is the indicator specifying the time period t^* at which the break occurs. Equation (15) is a break model that allows the intercept and the autoregressive coefficients to change.

The test devised by Andrews (1993) and Andrews & Ploberger (1994) uses a grid in searching for a single break occurring at the unknown date t^* with the best fit. Their test is a threshold autoregressive (TAR) model using t^* as the threshold variable. Note that the t- and F-statistics are biased because the searching algorithm uses the fixed time variable at each iteration, so that it is recommended that Hansen's (1999) bootstrapping procedure be used to obtain the critical values for the threshold model. The threshold model for a single break is given by

$$r\_ta_t = \begin{cases} \alpha_0 + \sum_{i=1}^{p} \alpha_i \, r\_ta_{t-i} + \varepsilon_{1t} \;\; if \; t > t^* \\ \gamma_0 + \sum_{i=1}^{p} \gamma_i r\_ta_{t-i} + \varepsilon_{2t} \;\; if t \leq t^* \end{cases} \tag{16}$$

where ε_{1t} and ε_{2t} indicate the heterogeneous residuals considered.

For estimating multiple breaks dates, Bai & Perron (2003) extend equation (16) to

$$r\_ta_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, r\_ta_{t-i} + \sum_{j=1}^{k} D_{jt} \left(\gamma_0 + \sum_{i=1}^{p} \gamma_i r\_ta_{t-i} \right) + \varepsilon_t \qquad (17)$$

where ε_t is homogeneous across the $k+1$ regimes. They recommend using a trimming value of 15%, that is, keeping at least 15% of the sample in a subsample, and setting the maximum number of breaks $k = 5$.

3.3.5.2 Results. Figure 7 plots the F-statistics calculated using equation (14). For simplicity, the AR (5) model is used instead of the ARMA (5, 2). The trimming value is set at 8% because of the relatively large sample size. A three-step procedure is recommended for two breaks:

Step 1 Search for the first breakpoint over the sample of 1-1600. As depicted in panel (a) of Figure 7, the F-statistic reaches its max at observation 143. The red line is the boundary at which the probability that the F-statistic exceeds the boundary value is $\alpha = 5\%$ under the null hypothesis of no structural change.

Step 2 Search for the second breakpoint over the sample of 144-1600. The F-statistic suggests that the break occurs at observation 1065 (see panel (b) of Figure 7).

Step 3 Use the determined second breakpoint as the end point of the sample and repeat Step 1, searching for the first break point. As presented in panel (c) of Figure 7, the break occurs at observation 138.

As suggested by the F-tests, the excess value series appear to have two structural breaks at 2 May 2014 and 9 January 2018.

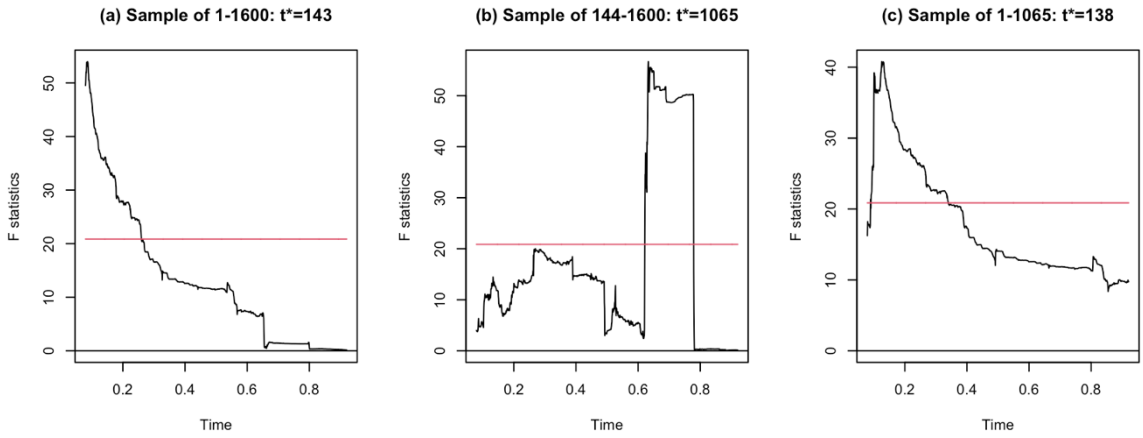

Figure 7. Plots of the F-statistics for the 3-step procedures: (a) on the sample of 1-1600, (b) on the sample of 144-1065, and (c) on the sample of 1-1065.

We now remove the MA terms and select a pure AR model. The best fit model is the AR (1) model given by

$$r\_ta_t = \underset{(0.001)}{-0.002} + \underset{(0.025)}{0.063} r\_ta_{t-1} + \varepsilon_t. \qquad (18)$$

This model has a tiny negative intercept and a small autoregressive coefficient, both statistically significant. The BIC is equal to -8069.

We then employ the Bai-Perron test on the AR (1) model using a 0.08 trimming and 5 breaks ($k = 5$). The model is specified in (17). The results are reported in Table 9.

Table 9. The results of the Bai-Perron test on the excess value series using 5 breaks.

Number of breaks	Break points at observation number				
$k = 1$	143				
$k = 2$	137				1065
$k = 3$	143			937	1065
$k = 4$	137		526	860	1065
$k = 5$	143	399	526	860	1065

Note: the test is based on the AR (1) model specified in equation (17).

The number of breaks is determined using the BIC criterion and the RSS. As depicted in Figure 8, the value of RSS decreases as the number of breaks is increasing, indicating the

more breaks the better fit for the model. However, the BIC starts to decrease after the first break, as adding one break will bring two parameters (one intercept and one slope coefficient) into the model, leading to a higher penalty of BIC scores. Though the BIC selected the model with one break, two breaks seems more realistic: the value of the BIC for two breaks is still lower than the value without a break and it has a lower RSS. Moreover, as shown in Figure 6, after January 2018, the series of the excess return behaves as a straight line except for occasional spikes. For 0, 1, and 2 break(s), the values of the BIC are -8064, -8097, and -8083, respectively.

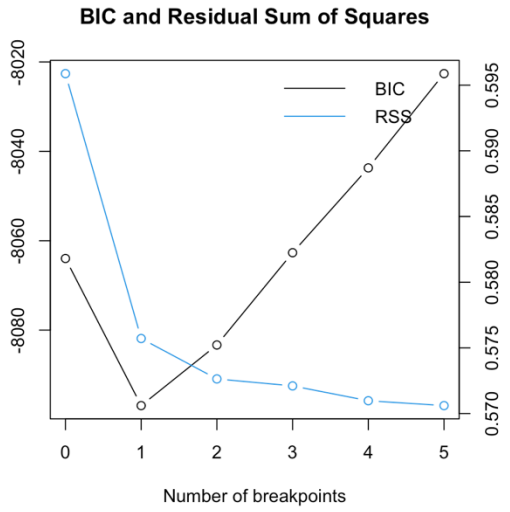

Figure 8. Plots of BIC (using the y-axis on the left) and residual sum of squares (RSS) (using the y-axis on the right).

In the case of two breaks, observation number 137 and 1965 indicate 1 April 2014 and 9 January 2018, respectively. The estimated model is given by

$$r_{ta_t} = \underset{(0.001)}{-0.010} - \underset{(0.027)}{0.006} r_{ta_{t-1}} + D_{1t}\left(\underset{(0.002)}{0.010} + \underset{(0.081)}{0.496} r_{ta_{t-1}}\right) + D_{2t}\left(\underset{(0.002)}{0.009} + \underset{(0.146)}{0.011} r_{ta_{t-1}}\right) + \varepsilon_t \quad (19)$$

where $D_{1t} = 1$ for $137 < t \leq 1065$, otherwise $D_{1t} = 0$; $D_{2t} = 1$ for $t > 1065$, otherwise $D_{2t} = 0$.

We can write equation (19) in the form:

$$r_{ta_t} = \begin{cases} -0.010 - 0.006 r_{ta_{t-1}} + \varepsilon_t & when\ t \leq 137 \\ 0.49 r_{ta_{t-1}} + \varepsilon_t & when\ \ 137 < t \leq 1065 \\ -0.001 + 0.005 r_{ta_{t-1}} + \varepsilon_t & when\ t > 1065 \end{cases} \quad (20)$$

Hence, by equation (20), the series of the excess value is split into three regimes:

Regime 1: From 11 September 2013 to 1 April 2014. The excess value series was extremely volatile, with very weak negative serial dependence. These results may indicate that the bitcoin exchange rate was not well-formed and thus oscillated widely about the FX spot.

Regime 2: From 2 April 2014 to 9 January 2018. The excess value series was less volatile, with relatively persistent positive serial dependence. The results may indicate that the bitcoin exchange rate was constructed with some bias against the FX spot, but the arbitrage chance gradually eliminated the discrepancy and even influenced market sentiment to the reverse side.

Regime 3: From 10 January 2018 to 6 March 2020. The excess value series had the smallest variations, with some spikes and very weak positive serial dependence. These results may indicate that the bitcoin exchange rate is locking the FX spot except for occasional malfunctions.

This chapter has described the statistical features and econometric time-series features of each individual time series. The level series of the bitcoin exchange rate, the bitcoin prices, the FX rates appear to follow a random walk process, while the return series, including the triangular arbitrage series, present a mean-reverting characteristic, suggesting a martingale. From this, we might infer a weak-form market efficiency regarding these prices or rates, although, strictly speaking, only the three FX rates coincide with this. Serial

autocorrelation over relatively long lags is observed with the bitcoin-related return series, and some of them present a nonzero mean. An ARMA (p, q) plus GARCH (1, 1) model using a student's t distribution is suggested for capturing the serial dependence. The estimated models appear to be adequate according to testing. The VaR statistics calculated from the empirical distributions suggest that the returns of the bitcoin exchange rate and the returns from the BX-FX triangular arbitrage have lower downward losses than the returns of the bitcoin prices at 1% and 2.5% probability levels but are higher than the returns of the FX rates. The tests for endogenous structural breaks detect two endogenous structural breaks existing in the triangular arbitrage series and, therefore, divide its sample period into three regimes, which raises the implication for the transformation of the bitcoin market in terms of currency trading.

4 Modeling I: In the long run

This chapter models the bitcoin exchange rate with other time series, focusing on the long-run relationship between them.

4.1 Long-run equilibrium

The law of one price reveals that identical goods sold in different locations should have the same price under the conditions of free competition, price flexibility, and no trade frictions. Both bitcoin markets and foreign exchange markets are thought of as decentralized markets, and they appear to stratify these conditions. There must be an equivalence of the USD/EUR bitcoin exchange rate that approximates the U.S. dollar price of a Euro in the bitcoin markets to the corresponding FX spot. As plotted in panel (c) of Figure 2, the series of the BX rate is intertwining with the series of the FX spot for the most time. The triangular arbitrage returns as the excess values between the BX rate and the FX spot present a mean-reverting feature. These facts all suggest that the BX rate and the FX spot are in equilibrium.

Hypothesis 4.1 There exists a long-run equilibrium between the USD/EUR bitcoin exchange rate and the corresponding FX spot rate.

4.1.1 OLS approach. The common approach to verify the long run correlationship between the BX rate and the FX spot is to use an OLS regression on the two series:

$$bx\_ue_t = \beta_0 + \beta_1 ue_t + \varepsilon_t. \tag{21}$$

The law of one price requires the coefficients $\beta_0 = 0$ and $\beta_1 = 1$. Coefficient β_0 is interpreted as a risk premium. A constant risk premium does not violate the one price assumption. There is a problem with this method: if regressors are unit-root processes, this regression is dismissed as 'spurious regression', leading to ineffective t- and F-statistics and biasedly estimated coefficients (Granger & Newbold, 1974). On the other hand, if the two unit-root processes are cointegrated of order (1, 1), an OLS regression yields a 'super-consistent' estimator of the co-integrating parameters β_0 and β_1 (Enders, 2014, p. 361).

The bitcoin exchange rate and the FX spot correspond to the case. Both series possess a unit-root. Nan & Kaizoji (2019b) find that the USD/EUR bitcoin exchange rate is cointegrated with the FX spot, futures, and short-period forwards rates over the period that falls into the second regime. Nevertheless, the OLS regression provides a fast and easy way to see the estimated coefficients.

4.1.1.1 Results. The estimated coefficients using OLS regression are reported in Table 10. The restrictions $\beta_0 = 0$ and $\beta_1 = 1$ appear to hold in Regime 2 and 3, but do not hold in Regime 1, which indicates the law of one price started to apply to the bitcoin markets only after Regime 1. No inference to the estimates can be conducted due to spurious regression. If the two series are cointegrated, these estimated values should be consistent and give the value of the cointegrating vector that makes the BX rate and the FX spot in the long-run equilibrium. A two-step methodology for testing cointegration is illustrated in the next section.

Table 10. The estimation of the long run equilibrium between the bitcoin exchange rate and the spot exchange rate using OLS regression.

Estimates	Regime 1 (137 obvs.)	Regime 2 (928 obvs.)	Regime 3 (535 obvs.)
$\hat{\beta}_0$	0.23	-0.00	-0.00
	(0.16)	(0.00)	(0.00)
$\hat{\beta}_1$	0.21	1.00	1.00
	(0.52)	(0.00)	(0.01)

Note: the standard errors are given in the parentheses.

4.1.2 Cointegration approach. Cointegration theory (Granger, 1986) says that for two series both integrated of order one, denoted $I(1)$, if there exists a linear combination of them to become an $I(0)$ process, the two series are said to be cointegrated of order (1, 1), denoted $CI(1,1)$ (Enders, 2014).

Generally, let $\boldsymbol{y}_t$ be a $n \times 1$ vector of time series that are all $I(1)$ and $\boldsymbol{\beta}$ be an $n \times 1$ vector. If $\boldsymbol{\beta}$ imposes the linear restriction

$$\boldsymbol{\beta}'\boldsymbol{y}_t = 0 \tag{22}$$

then $\boldsymbol{y}_t$ is said to be in equilibrium; a condition for this is defined as co-integration. Vector $\boldsymbol{\beta}$ is then called the co-integrating vector. Most of the time, $\boldsymbol{y}_t$ is not in equilibrium. Let

$$z_t = \boldsymbol{\beta}'\boldsymbol{y}_t \tag{23}$$

which is a stationary process, and call quantity z_t the equilibrium error (Granger, 1986).

Let $\boldsymbol{y}_t = (1, ue_t, bx\_ue_t)'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$. The presence of co-integration indicates that there exists a vector $\boldsymbol{\beta}$ in which β_2 is normalized to minus unity produces

$$\boldsymbol{\beta}'\boldsymbol{y}_t = (\beta_0 + \beta_1 ue_t - bx\_ue_t) = 0 \tag{24}$$

Rewriting (23), we have

$$bx\_ue_t = \beta_0 + \beta_1 ue_t \tag{25}$$

which is equivalent to (21).

Engle & Granger (1987) propose a two-step methodology to perform co-integration analysis: First, OLS regression is used to produce the “super consistent” coefficients of the

co-integrating vector like what we have done in section 4.1.1, then the equilibrium error formed by (23) for stationarity is tested. However, the methodology is posited on a lack of accuracy, as errors will be accumulated through the two-step procedure and there are no statistics for directly testing for co-integration.

Johansen (1988) proposes a statistical approach to test the hypothesis of co-integration among nonstationary variables, and appropriate statistics are derived from the maximum likelihood estimation involved.

We follow the notation that Johansen and Juselius use to illustrate the procedure. Consider a vector autoregressive model (VAR) that treats all variables as endogenous:

$$H_0: \boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Pi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Pi}_k \boldsymbol{y}_{t-k} + \boldsymbol{\varepsilon}_t \quad (26)$$

where $\boldsymbol{y}_t$ denotes a p-dimensional vector of economic variables, $\boldsymbol{\mu}$ denotes a constant term and $\boldsymbol{\varepsilon}_t$ denotes a p-dimensional vector of error terms with an i.i.d. Gaussian distribution. The $p \times p$ matrices $\boldsymbol{\Pi}_1, \ldots \boldsymbol{\Pi}_k$ are coefficients for k-lags of $\boldsymbol{y}_t$ and $t = 1, \ldots, T$. The unrestricted VAR model has Tp observations and $p + kp^2 + p(p+1)/2$ parameters (here, $p(p+1)/2$ comes from the symmetric variance-covariance matrix). The null hypothesis, denoted H_0, assumes absence of co-integration among the variables in vector $\boldsymbol{y}_t$.

Add and subtract $\boldsymbol{\Pi}_k \boldsymbol{y}_{t-k+1}$ to the right-hand side of (25) to obtain

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Pi}_1 \boldsymbol{y}_{t-1} + \cdots + (\boldsymbol{\Pi}_{k-1} + \boldsymbol{\Pi}_k) \boldsymbol{y}_{t-k+1} - \boldsymbol{\Pi}_k \Delta \boldsymbol{y}_{t-k+1} + \boldsymbol{\varepsilon}_t \quad (27)$$

Then, add and subtract $(\boldsymbol{\Pi}_{k-1} + \boldsymbol{\Pi}_k) \boldsymbol{y}_{t-k+2}$ and continue in this fashion to obtain

$$\boldsymbol{y}_t = \boldsymbol{\mu} + (\boldsymbol{\Pi}_1 + \cdots + \boldsymbol{\Pi}_{\mathrm{k}}) \boldsymbol{y}_{t-1} - (\boldsymbol{\Pi}_2 + \cdots + \boldsymbol{\Pi}_{\mathrm{k}}) \Delta \boldsymbol{y}_{t-1} - \cdots - \boldsymbol{\Pi}_k \Delta \boldsymbol{y}_{t-k+1} + \boldsymbol{\varepsilon}_t \quad (28)$$

Next, subtract $\boldsymbol{y}_{t-1}$ from both sides of the equation above and rewrite it as

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} \ldots + \boldsymbol{\Gamma}_{\mathrm{k}-1} \Delta \boldsymbol{y}_{t-k+1} + \boldsymbol{\varepsilon}_t \quad (29)$$

where

$$\Delta \boldsymbol{y}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t-1}$$

$$\boldsymbol{\Gamma}_{\mathrm{k}-1} = -\boldsymbol{\Pi}_k$$

$$\boldsymbol{\Gamma}_{\mathrm{i}} = -(\boldsymbol{\Pi}_{i+1} + \cdots + \boldsymbol{\Pi}_k), \qquad (i = 1, \ldots, k-2)$$

and

$$\boldsymbol{\Pi} = (\boldsymbol{\Pi}_1 + \cdots + \boldsymbol{\Pi}_{\mathrm{k}} - \mathbf{I})$$

At first glance, equation (29) appears to be a first-order differenced VAR model except for the term $\boldsymbol{\Pi}\boldsymbol{y}_{t-1}$, or it might possibly be thought to be a vector version of the ADF test. However, neither of these impressions is actually true. The key point is that the coefficient matrix $\boldsymbol{\Pi}$ is of special interest—the information about co-integration.

The key feature to note is the rank of the p-dimensional coefficient matrix $\boldsymbol{\Pi}$, which is also called the transitory impact matrix.

There are three cases related to the rank of this matrix:

- Case (I): $rank(\boldsymbol{\Pi}) = p$, or matrix $\boldsymbol{\Pi}$ has full rank, indicating that all variables in vector $\boldsymbol{y}_t$ are stationary; since there are p linear restrictions on p variables, $\boldsymbol{\Pi}\boldsymbol{y}_{t-1}$ can only be stationary so that (29) holds. The model in (26) is appropriate, as all variables are stationary.
- Case (II): $rank(\boldsymbol{\Pi}) = 0$ or matrix $\boldsymbol{\Pi}$ is null, suggesting that all p variables are non-stationary and there is no linear restriction on $\boldsymbol{y}_{t-1}$ to make it stationary, i.e. there is no co-integration among the variables in vector $\boldsymbol{y}_t$. Equation (29) becomes a first-order differenced VAR with nonstationary variables that are all $I(1)$.
- Case (III): $\mathrm{rank}(\boldsymbol{\Pi}) = r$, where $r < p$, indicates that r restrictions could make $\boldsymbol{\Pi}\boldsymbol{y}_{t-1}$ stationary; hence, the rank of matrix $\boldsymbol{\Pi}$ is equal to the number of distinct co-integrating vectors. The term $\boldsymbol{\Pi}\boldsymbol{y}_{t-1}$ becomes the error-correcting term and (29) is a vector error correcting model (VECM).

In linear algebra, the rank of a matrix is equal to the number of nonzero characteristic roots; Eigenvalues are used solving this problem. Johansen, however, designs the specific

eigenvalues to be non-negative and real and to represent linear combinations of the data that have maximum canonical correlations. Based on maximum likelihood estimation, his two likelihood ratio test statistics—Q-statistics—are:

(i) The trace statistic (denoted Q_{trace}):

$$-2\ln(Q; H_1|H_0) = -T\sum_{i=r+1}^{p}\ln(1-\hat{\lambda}_i) \tag{30}$$

which tests for whether the rank of matrix $\mathbf{\Pi}$ is equal to r, where $r < p$. The quantity $-2\ln(\cdot)$ is the logarithmic likelihood ratio and $H_1|H_0$ denotes that the statistic is for hypothesis H_1 in H_0, as H_0 is a special case of H_1 when $r = p$ as described in Case (I). $\hat{\lambda}_i$ ($i = 1, \dots, p$) is the set of estimated eigenvalues, where $\hat{\lambda}_i \in \mathbb{R}$ and $\hat{\lambda}_1 > \cdots > \hat{\lambda}_p \geq 0$. The condition that $rank(\mathbf{\Pi}) = r$ is equivalent to the number of non-zero $\hat{\lambda}_i$ being r; hence, the statistic tests whether the summation of the remaining $(p - r)$ engenvalues, $\hat{\lambda}_{r+1}, \dots, \hat{\lambda}_p$, is equal to zero. When the value of $\hat{\lambda}_{r+1}$ is very close to zero, $\ln(1 - \hat{\lambda}_{r+1})$ is also close to zero, so that $-T\sum_{i=r+1}^{p}\ln(1-\hat{\lambda}_i)$ is close to zero; otherwise, the value of $-T\sum_{i=r+1}^{p}\ln(1-\hat{\lambda}_i)$ will be negative.

(ii) The maximum eigenvalue statistic (denoted Q_{max}):

$$-2\ln(Q; r|r+1) = -T\ln(1-\hat{\lambda}_{r+1}) \tag{31}$$

which examines whether the r-th eigenvalue is nonzero relative to the alternative that the $(r + 1)$-th eigenvalue is nonzero. For instance, we test the absence of co-integration, or $r = 0$, null hypothesis against the alternative hypothesis positing the presence of an independent nonzero eigenvalue, i.e. $\hat{\lambda}_1$, the largest eigenvalue is not equal to zero, or $r = 1$. If the value of $\hat{\lambda}_1$ is significantly different from zero, the value of $-T\ln(1 - \hat{\lambda}_{r+1})$ is negative and its absolute value should be greater than a given critical value, suggesting rejection of the null

hypothesis. Hence, we can say that we reject the null of $r = 0$ and tend toward accepting the alternative of $r = 1$ at some significance level, for instance 5%.

4.1.2.1 Results. Table 11 reports the results of lag selection, the Johansen tests, and the model adequacy tests. For lag selections, the procedure begins by testing an unrestricted VAR model using data in level. The lag length is first suggested by the AIC and the final prediction error (FPE) criteria as they prefer to incorporate more lags so that more information is included in the autoregressive terms. The portmanteau statistic is used for testing the absence of up to the order 16 serially correlated disturbance of (26). The time-series vector includes three entries: $\boldsymbol{y}_t = (1, ue_t, bx\_ue_t)'$. The AIC and FPE criteria have identical results: suggesting that 1, 3, and 1 lag(s) should be incorporated in the unrestricted VAR models of Regime 1-3, respectively. The Portmanteau-test is suggestive of the adequacy of the three VAR models.

Table 11. The results of lag selection, the Johansen tests, and the model adequacy diagnostics

	$\boldsymbol{y}_t = (1, ue_t, bx\_ue_t)'$		
	Regime 1 (137 obvs.)	**Regime 2 (928 obvs.)**	**Regime 3 (535 obvs.)**
$Lags_{var}$	1	3	1
$P_{var}(16)$	64.86	57.61	55.21
$Lags_{joh}$	2	3	2
$Q_{trace}(H_0: r = 0)$	72.94***	120.91***	225.66***
$Q_{trace}(H_0: r \leq 1)$	9.22*	6.10	3.24
$Q_{max}(H_0: r = 0)$	63.72***	114.81***	222.42***
$Q_{max}(H_0: r = 1)$	9.22*	6.10	3.24
$\hat{\beta}_0$	0.221	-0.000	-0.001
$\hat{\beta}_1$	0.258	1.000	1.002
$P_{joh}(16)$	62.85	57.09	52.61

Note: $Lags_{var}$ denotes the lag length suggested by the AIC and FPE criteria using the unrestricted VAR model with the level series. $P_{var}(16)$ denotes the Portmanteau-statistic testing on 16 lags of the disturbances the VAR model for serial autocorrelation. Q_{trace} and Q_{max} statistics are Johansen's statistics testing for the number of nonzero eigenvalues and the marginal nonzero eigenvalue, respectively; the null hypothesis is listed in the parentheses. β_0 and β_1 are the entries in the cointegrating vector in (24). $P_{joh}(16)$ denotes the Portmanteau-statistic for the serially correlated disturbances of the Johansen test.

*, **, and *** significant at 10%, 5%, and 1%, respectively.

The results of the Johansen tests for Regime 1 suggest the existence of one cointegrating vector for $\boldsymbol{y}_t$. Both Q_{trace} and Q_{max} statistics reject the null hypothesis of $rank(\boldsymbol{\Pi}) = r = 0$, i.e., the absence of the cointegrating vector at the 1% significance level. The inferences on $r \leq 1$ are significant at 10%, indicating, by 10% chances, bx_ue_t and ue_t are both stationary in Regime 1. $P_{joh}(16)$ is suggestive of adequacy of the model.

The estimated $\hat{\beta}_0$ and $\hat{\beta}_1$ give the specification of

$$0.221 + 0.258ue_t - bx\_ue_t = 0$$

or

$$bx\_ue_t = 0.221 + 0.258ue_t. \tag{32}$$

Equation (32) suggests the linear specification of the long-run equilibrium between bx_ue_t and ue_t in Regime 1. The result raises the implication of 'different prices' between bitcoin markets and the FX market. As shown in panel (a) of Figure 9, the equivalence of the two logarithmic prices occurs at 0.2978, indicating the two prices are equivalent at $e^{0.2978} = 1.347$. When the USD/EUR rate is greater than 1.347, $bx\_ue_t < ue_t$, indicating depreciated Euros in bitcoin markets, whereas when the USD/EUR rate is less than 1.347, $bx\_ue_t > ue_t$, indicating appreciated Euros in bitcoin markets.

This may be because that the value of a Euro kept appreciating over the period of Regime 1 (see panel (b) of Figure 9) so that European investors have more propensity to buy bitcoins than the Americans. Also, the bitcoin markets were not mature to response the information from the FX markets, so the violation of the law of one price did not be corrected by arbitrage. The result is coincide with Dong & Dong's (2015) findings that the persistent discrepancies may result from the buy-and-hold strategy that performed by the investors during the period of Regime 1.

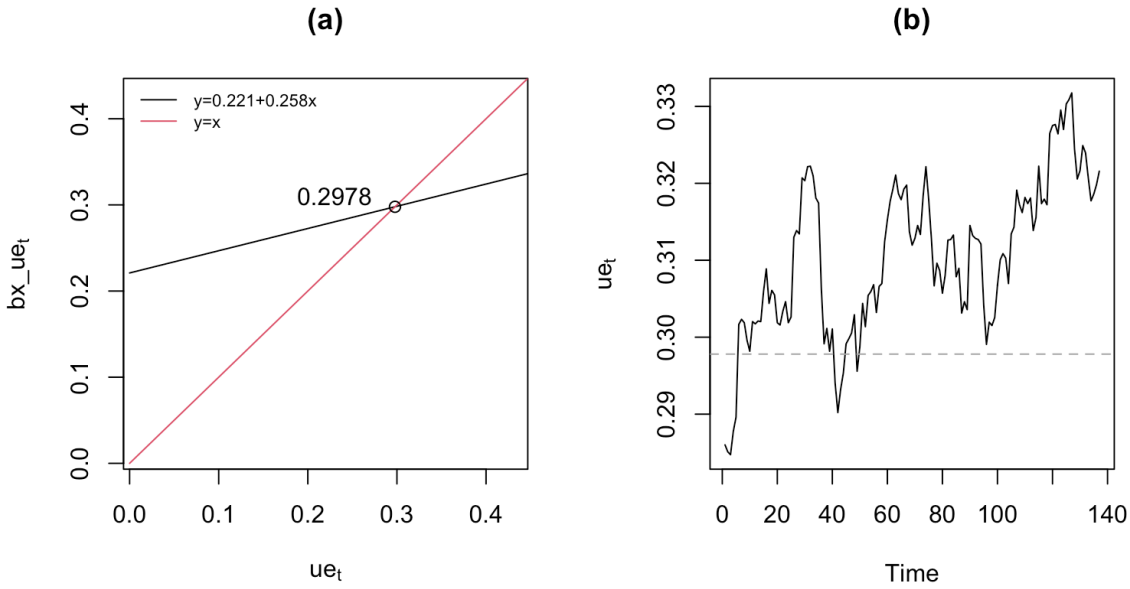

Figure 9. The plots of (a) the linear specification of the equilibrium between the USD/EUR bitcoin exchange rate and the FX spot and (b) the USD/EUR FX spot series in Regime 1. Note: All rates are in natural logarithm.

For Regime 2, one cointegrating vector is found. The estimated $\hat{\beta}_0$ and $\hat{\beta}_1$ give the specification of

$$0 + ue_t - bx\_ue_t = 0$$

or

$$bx\_ue_t = ue_t$$

indicating that the law of one price, the USD/EUR rate, is held in both the bitcoin markets and the FX spot markets. The $P_{joh}(16)$ statistic is suggestive of adequacy.

For Regime 3, the Johansen test suggests one significant cointegrating vector. The estimated $\hat{\beta}_0$ and $\hat{\beta}_1$ give the specification of

$$-0.001 + 1.002ue_t - bx\_ue_t = 0$$

or

$$bx\_ue_t = 1.002ue_t - 0.001.$$

The $P_{joh}(16)$ statistic is suggestive of adequacy.

4.2 Market efficiency in the semi-strong form and the unbiased estimator

The primary role of a bitcoin market is to provide an online "gathering" of participants for the purchase and sale of provisions of bitcoin. Because of the ease of bitcoin transferring and low costs, bitcoins can move from market to market, so different currencies are connected by bitcoins. For bitcoin markets, foreign exchange rates serve as a piece of publicly available information. However, if bitcoin markets are not so efficient such that the information of the foreign exchange rate is not reflected by the prices of bitcoin, a well-informed speculator will arbitrage on the price discrepancy.

For example, a bitcoin is traded for $10,000 in an American bitcoin market, and the USD/EUR rate is quoted at 1.18. Assume bitcoin's attractiveness is the most significant factor to its price formation, and European investors are less interested in buying bitcoins, so they would only like to spend €8,000 for a bitcoin. As such, our proposed bitcoin exchange rate is at 10000/8000=1.25, indicating appreciated Euros in the bitcoin market. A shrewd speculator will use, say, €80,000 to buy 10 bitcoins in the European bitcoin market, and trade the 10 bitcoins in the American market for $100,000, and then exchange $100,000 into about €84746 at 1.18 USD/EUR in the FX market. The return is about (84746-80000)/80000=5.9%. Before long, the arbitrage will result in the appreciation of bitcoin in the European market. This example illustrates that an efficient market should also include publicly available information into prices.

The semi-strong form of market efficiency emphasizes that prices reflect publicly available information, so the fundamental analysis is useless for the current price (Fama, 1970). Our semi-strong form test is presented by

$$E(bx\_ue_t|\Phi_{t-1}) = ue_t \tag{33}$$

where Φ_{t-1} is one-period-ahead publicly available information from both the FX and bitcoin markets. It says that the current expected bitcoin exchange rate conditioned to the historical information set, Φ_{t-1}, is equal to the current FX spot rate.

The excess market value between the bitcoin and FX exchange rates, denoted x_t, is expressed by

$$x_t = E(bx\_ue_t|\Phi_{t-1}) - ue_t. \tag{34}$$

Then, the conditional expectation of the excess market value has a zero mean:

$$\begin{aligned} E(x_t|\Phi_{t-1}) &= E[E(bx\_ue_t|\Phi_{t-1})|\Phi_{t-1}] - E(ue_t|\Phi_{t-1}) \\ &= E(ue_t|\Phi_{t-1}) - E(ue_t|\Phi_{t-1}) = 0 \end{aligned} \tag{35}$$

which says that the sequence $\{x_t\}$ is a 'fair game' with respect to information sequence $\{\Phi_{t-1}\}$, or, equivalently, $\{x_t\}$ is a martingale.

Hakkio & Rush (1989) propose the joint equilibrium conditions: (i) no risk premium and (ii) the rational use of available information , the equivalence to the constraints $\beta_0 = 0$ and $\beta_1 = 1$ in (21).

In presence of cointegration, these restrictions ($\beta_0 = 0$ and $\beta_1 = 1$) on the cointegrating vector give

$$z_t = \boldsymbol{\beta}'\boldsymbol{y}_t = (\beta_0 + \beta_1 ue_t - bx\_ue_t) = ue_t - bx\_ue_t \tag{36}$$

where z_t denotes the equilibrium error in (23).

Then

$$E(z_t|\Phi_{t-1}) = 0 \tag{37}$$

so that the equilibrium error sequence $\{z_t\}$ is a fair game as well. Note that $E(z_t|\Phi_{t-1}) = E(-x_t|\Phi_{t-1})$.

In this way, the semi-strong test for market efficiency is connected with the cointegration test; the restrictions $\beta_0 = 0$ and $\beta_1 = 1$ are saying that the current FX spot is an unbiased estimator to the current bitcoin exchange rate, conditioned to the historical and publicly available information set Φ_{t-1}.

Hypothesis 4.2 the current FX spot is an unbiased estimator to the current bitcoin exchange rate, conditioned to the historical and publicly available information set Φ_{t-1}.

4.2.1 Methodology. In the Johansen test, matrix $\mathbf{\Pi}$ can be decomposed into vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. During the process of identifying the rank of matrix $\mathbf{\Pi}$, the eigenvalues $\hat{\lambda}_1 > \cdots > \hat{\lambda}_p$ and the relative normalized eigenvectors $\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_p$ are estimated to represent the linear combinations of the data that have maximum canonical correlations. That $rank(\mathbf{\Pi}) = r$, where $0 < r < p$, indicates that co-integration is present and that the number of independent co-integrating vectors is equal to r. If we choose matrix $\widehat{\boldsymbol{\beta}}$ as

$$\widehat{\boldsymbol{\beta}} = (\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_r)' \tag{38}$$

and let $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, then the estimated matrix $\widehat{\boldsymbol{\alpha}}$ is calculated by $\widehat{\mathbf{\Pi}}$ and $\widehat{\boldsymbol{\beta}}$. Matrix $\widehat{\boldsymbol{\beta}}$ gives the co-integrating vectors, since $\widehat{\boldsymbol{\beta}}$ is chosen from the effective linear combinations of data that have maximum canonical correlations, indicative of long-run equilibrium; matrix $\widehat{\boldsymbol{\alpha}}$ gives the speed-of-adjustment coefficient, which we will discuss it in the next chapter.

Thus, the hypothesis of cointegration is equivalent to $\mathbf{\Pi}$ having a reduced rank form and being able to be decomposed into $\boldsymbol{\alpha}\boldsymbol{\beta}'$. This situation is represented as

$$H_1: \mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' \tag{39}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have dimension of $p \times r$. The hypothesis H_1, asserting the presence of co-integration, can be tested towards H_0 as shown in (30). The marginal eigenvalue test for $\hat{\lambda}_r$ against $\hat{\lambda}_{r+1}$ is represented in (31).

Similarly, with respect to the presence of co-integration expressed in H_1, linear restrictions can be placed on either $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ in the inference process. Following the fashion of $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, further decompositions can be performed as $\boldsymbol{\alpha} = \boldsymbol{A}\boldsymbol{\psi}$ and $\boldsymbol{\beta} = \boldsymbol{H}\boldsymbol{\varphi}$, with matrix dimensions $\boldsymbol{A}\ (p \times m)$, $\boldsymbol{\psi}\ (m \times r)$, $\boldsymbol{H}\ (p \times s)$ and $\boldsymbol{\varphi}\ (s \times r)$, where $r \leq s \leq p$ and $r \leq m \leq p$. Linear restrictions can be placed on parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by constraining matrices $\boldsymbol{A}$ and $\boldsymbol{H}$. The restrictions reduce the parameters from $\boldsymbol{\alpha}\ (p \times r)$ and $\boldsymbol{\beta}\ (p \times r)$ to $\boldsymbol{\psi}\ (m \times r)$ and $\boldsymbol{\varphi}\ (s \times r)$. By constraining metrics $\boldsymbol{H}$, we can place linear restrictions on $\boldsymbol{\beta}$.

Here, we describe the inference process:

$$H_2: \boldsymbol{\beta} = \boldsymbol{H}\boldsymbol{\varphi} \tag{40}$$

Note that $H_2 \subset H_1$ meaning the H_2 hypothesis is a special case of H_1. The H_0 hypothesis, where matrix $\boldsymbol{\Pi}$ is unrestricted, is also a special case of H_1 when $r = p$.

For H_2, we solve the specific formula for eigenvalues $\hat{\lambda}_{2.1} > \cdots > \hat{\lambda}_{2.s}$ and the normalized eigenvectors $\hat{\boldsymbol{v}}_{2.1}, \ldots, \hat{\boldsymbol{v}}_{2.s}$, where $r \leq s \leq p$. Choose $\hat{\boldsymbol{\varphi}} = (\hat{\boldsymbol{v}}_{2.1}, \ldots, \hat{\boldsymbol{v}}_{2.r})'$ and $\hat{\boldsymbol{\beta}} = \boldsymbol{H}\hat{\boldsymbol{\varphi}}$ and find $\hat{\boldsymbol{\alpha}}$ by solving $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$. The approximate likelihood ratio test of hypothesis H_2 in H_1 is given by

$$-2\ln(Q; H_2|H_1) = T\sum_{i=1}^{r} \ln\{(1-\hat{\lambda}_{2.i})/(1-\hat{\lambda}_i)\} \tag{41}$$

where $\hat{\lambda}_i, i = 1, \ldots, r$ are eigenvalues under H_1. The asymptotic distribution of this statistic is given by χ^2 with $r(p-s)$ degrees of freedom (Johansen, 1991).

In this thesis, the cointegrating vector is specified as $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ where β_2 is normalized to -1. The restriction $\beta_1 = 1$ is equivalent to $\beta_1 = -\beta_2$. Hypothesis $H_{2.1}$: $\hat{\beta}_1 = -\hat{\beta}_2$ is realized by constraining $\boldsymbol{H}$ as

$$\boldsymbol{H} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ -1 & 0 \end{pmatrix}. \tag{42}$$

Hypothesis $H_{2.2}$: $\hat{\beta}_0 = 0$ is realized by constraining $\boldsymbol{H}$ as

$$\boldsymbol{H} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{43}$$

4.2.2 Results. The law of one price appear not to hold for Regime 1. The inferences on estimated $\widehat{\boldsymbol{\beta}}$ are conducted for Regime 2 and 3. Table 12 reports the results. For $H_{2.1}$ and $H_{2.2}$, the Q-statistics follow the χ^2 distribution with degrees of freedom of one. For $H_{2.3}$, the asymptotic distribution of this statistic is given by χ^2 with 2 degrees of freedom. In Regime 2, all three hypotheses $H_{2.i}$, for $i = 1,2,3$, are not rejected. These results indicate that the restrictions $\beta_0 = 0$ and $\beta_1 = 1$ hold for the period of Regime 2. For Regime 3, $\beta_0 = 0$ or $\beta_1 = 1$ appear to hold individually, but the Q-statistic rejects $H_{2.3}$ at 1%, indicating the two restrictions do not jointly hold. This may result from that when we restrict $\beta_1 = 1$, $\hat{\beta}_0 = 0.0011$ and when we restrict $\beta_0 = 0$, $\hat{\beta}_1 = 0.9923$.

Table 12. Inferences on $\widehat{\boldsymbol{\beta}}$ that is estimated from the Johansen tests over Regime 2 and Regime 3.

	$\boldsymbol{H_1}$ $\boldsymbol{\Pi = \alpha\beta'}$	$\boldsymbol{H_{2.1}}$ $\hat{\beta}_1 = -\hat{\beta}_2$	$\boldsymbol{H_{2.2}}$ $\hat{\beta}_0 = 0$	$\boldsymbol{H_{2.3}}$ $\hat{\beta}_1 = -\hat{\beta}_2$ and $\hat{\beta}_0 = 0$
	Regime 2			
$Q(H_{2.i}\|H_1)$		0.01	0.22	1.48
$\hat{\beta}_0$	-0.000	0.000	0.000	0.000
$\hat{\beta}_1$	1.000	1.000	1.000	1.000
$\hat{\beta}_2$	-1.000	-1.000	-1.000	-1.000

	Regime 3			
$Q(H_{2.i}\|H_1)$		0.07	2.03	19.34***
$\hat{\beta}_0$	-0.001	-0.001	0.000	0.00
$\hat{\beta}_1$	1.002	1.000	0.992	1.00
$\hat{\beta}_2$	-1.000	-1.000	-1.000	-1.00

Note: $Q(H_{2.i}|H_1)$ denotes the Q statistic for the likelihood ratio test of (41). The null hypotheses $H_{2.i}$, for $i = 1, 2, 3$, specified in row 2 of this table, is tested against the alternative hypothesis H_1, the presence of one cointegrating vector. $\widehat{\boldsymbol{\beta}}$ is normalized so that $\hat{\beta}_2 = -1$.
*** significant at 1%.

In conclusion, the FX spot is an unbiased estimator to the BX rate over Regime 2 and 3, but there exists a slight negative risk premium over Regime 3. Because of this unbiasedness, the bitcoin markets are considered to follow the semi-strong form efficient market hypothesis with respect to the FX spot market in the long run.

4.3 Testing for covered interest parity

Covered interest parity (CIP) is commonly used for testing market efficiency. The CIP posits that the covered (or the hedged) return from investment in a foreign currency should be equal to the return from investment in the domestic currency, regardless of the level of the two interest rates (Crowder, 1995):

$$\frac{1 + i_{d,t}}{1 + i_{f,t}} = \frac{f_t^m}{s_t} \tag{44}$$

where $i_{d,t}$ denotes the interest rate in the domestic currency, $i_{f,t}$ denotes the interest rate in the foreign currency, s_t denotes the current spot exchange rate, and f_t^m denotes the forward foreign exchange rate with maturity horizon m. The CIP relation contains two conditions associated with market efficiency: (i) risk neutrality, and (ii) rational expectations, so that the FX forward rate is an unbiased predictor of the future FX spot rate (Kang, 2019), as given by

$$E_t(s_{t+m}) = f_t^m. \tag{45}$$

Under these two assumptions, the CIP suggests a certain level of market efficiency, as violations of the CIP result in risk-free arbitrage opportunities to speculators. If these opportunities are outside the transaction-cost band, an efficient market will eliminate the deviations.

Many studies have found evidence of CIP violations. However, the existence of profitable CIP deviations "shed no light on the empirical validity of the efficiency hypothesis" (Clinton, 1988), and only persistent profitable deviations violate the EMH. The arbitrage paradox states that the market is efficient, yet a short-run arbitrage opportunity is simultaneously created when investors may not have sufficient incentives to observe the market (Grossman & Stiglitz, 1980; Akram et al., 2008). Kang (2019) proposes that when the two assumptions do not hold, testing for the EMH may not be valid, i.e., rejection does not guarantee the violation of the EMH. Crowder (1995) argues that persistent riskless profits can be reconciled with market efficiency.

Hypothesis 4.3 The bitcoin exchange rate of USD/EUR coincides with the covered interest rate parity.

4.3.1 Methodology. The test for the CIP relation is based on daily data. The data consist of the USD/EUR bitcoin exchange rate, the one-month forward foreign exchange rate, and the one-month LIBOR ICE deposit rates for the U.S. dollar and Euro. To maintain the correct sequence of transactions, we use the bid prices of the deposit rates. Specifically, a speculator can either lend dollars at the domestic market or buy Euros at the FX market, and then lend Euros in the European market, simultaneously hedging its Euro position on the forward market (selling Euros forward).

We take logarithm of equation (44) to obtain

$$\log\left(1+\frac{i_{d,t}}{12}\right)-\log\left(1+\frac{i_{f,t}}{12}\right)=\log(f_t^m)-\log(s_t). \tag{46}$$

where $i_{d,t}/12$ denotes the one-month LIBOR ICE deposit rate for the U.S. dollar (the domestic) and $i_{f,t}/12$ denotes the one-month LIBOR ICE deposit rate for the Euro (the foreign)[1].

[1]I thank professor Kaneko for helping me point out the mistake that I didn't divide the annualized one-month deposit rate by 12, which approximates the deposit rate for one month.

The left-hand side of (46) refers to the interest differential, denoted id_t; the right-hand side of (46) refers to the forward premium, denoted fp_t. The forward-BX differential, denoted fp_bx_t, is constructed using $fo\_ue_t - bx\_ue_t$ (the series are in logarithm form). The CIP relation is then investigated as a cointegration relation using the Johansen test.

4.3.2 Results. Figure 10 plots the three time-series of the interest differential, the FX forward premium, and the forward premium on the BX rate. The series of the interest differential and the forward premium present a similar pattern but are different in scales. The series forward rate on the BX rate behaves more like the triangular arbitrage series. The statistics of the ADF and KPSS tests suggest that the interest differential and forward premium contain unit roots, while the forward premium on the BX rate appears to be stationary.

We first investigate the CIP relation in the FX market. The Johansen test rejects $H_0: r = 0$ at the 1% significance level but does not reject $H_0: r \leq 1$. Based on the estimated coefficients, the error-correction model (ECM) representation is given by

$$\Delta fp_t = -0.081(fp_{t-1} - 1.057id_{t-1}) + \sum_{i=1}^{7}\gamma_{1i}\Delta fp_{t-i} + \varepsilon_{1t} \tag{47}$$

$$\Delta id_t = 0.001(fp_{t-1} - 1.057id_{t-1}) + \sum_{i=1}^{7}\gamma_{2i}\Delta id_{t-i} + \varepsilon_{2t} \tag{48}$$

This model suggests that the long-run equilibrium relation is $fp_{t-1} = 1.0057id_{t-1}$, so the CIP holds in the long run. In the short run, when a deviation from the equilibrium relation suggested by $fp_{t-1} - 1.057id_{t-1}$ occurs one day before, the current change in the premium forward will decrease by 8.1% of the deviation and the current change in the interest differential will increases by 0.1% of the deviation, moving together to eliminate the discrepancy.

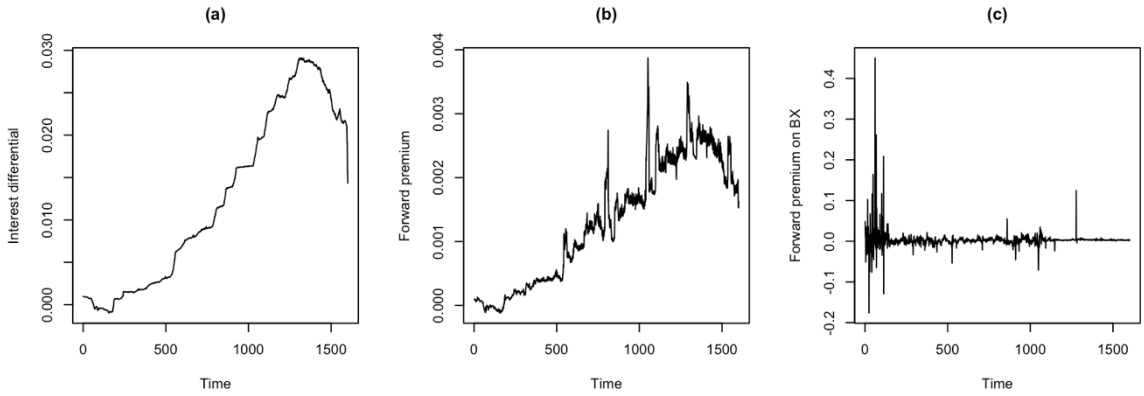

Figure 10. Series plots: (a) interest differential between the one-month LIBOR deposit rates of US dollars and Euros, (b) forward premium, and (c) forward premium on the bitcoin exchange rate.

We next investigate whether the CIP relation holds regarding the bitcoin exchange rate. The results suggest that the forward-BX differential is cointegrated with the interest differential over the Regime period. Over the Regime 2 period, two cointegrating vectors are found, indicating the two series are stationary.

For Regime 3, the Johansen test gives

$$\Delta fp\_bx_t = -0.995(fp\_bx_{t-1} - 2.096id_{t-1} + 0.001) + \sum_{i=1}^{4} \gamma_{1i}\Delta fp\_bx_{t-i} + \varepsilon_{1t} \quad (49)$$

$$\Delta id_t = -0.000(fp\_bx_{t-1} - 2.096id_{t-1} + 0.001) + \sum_{i=1}^{4} \gamma_{2i}\Delta id_{t-i} + \varepsilon_{2t}.$$

The long-run equilibrium is represented by $fp\_bx_{t-1} = 2.096id_{t-1} - 0.001$. Deviations from the CIP relation are adjusted singly by the changes in forward premium on the BX rate: Δfp_bx_t decreases it by 99.5% of the discrepancy. The series of id_t and fp_bx_t are plotted in panels (a) and (b) of Figure 11, respectively. There are several spikes in the series of fp_bx_t during the Regime 3 period.

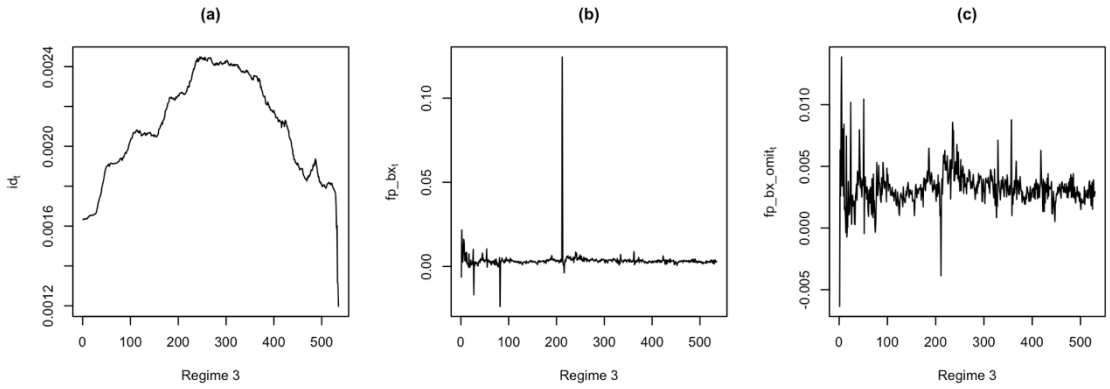

Figure 11. The plots of the series over Regime 3: (a) the interest differential, (b) the forward premium on the BX rate, and (c) the forward premium on the BX rate omitting 5 outliers whose absolute values are greater than 0.015.

Consider removing the absolute values of fp_bx_t greater than 0.015 so that five outliers are omitted. The omitted fp_bx_t series, denoted $fp\_bx\_omit_t$, is plotted in panel (c) of Figure 11. Though the behavior of $fp\_bx\_omit_t$ is still volatile, it presents a concave pattern, as does id_t. For $fp\_bx\_omit_t$, the unit-root tests provide controversial results: The models with a trend term suggest stationarity, while the model with a constant suggests unit roots. This ambiguity indicates that $fp\_bx\_omit_t$ may be either a difference-stationary (DS) series or a trend-stationary (TS) series. The estimated Johansen test gives

$$\Delta fp\_bx_t = -0.536(fp\_bx_{t-1} - 1.706id_{t-1}) + \sum_{i=1}^{6} \gamma_{1i}\Delta fp\_bx_{t-i} + \varepsilon_{1t}$$
$$\Delta id_t = -0.000(fp\_bx_{t-1} - 1.706id_{t-1}) + \sum_{i=1}^{6} \gamma_{2i}\Delta id_{t-i} + \varepsilon_{2t}. \quad (50)$$

The model suggests a different long-run equilibrium relation $fp\_bx_{t-1} = 1.706id_{t-1}$ and a relatively slow adjustment process, as indicated by -0.536.

4.4 An attractor

In economic theories, an attractor is the long-run equilibrium value to which a stationary time series reverts. For a stationary series of $\{y_t\}$, the Dickey & Fuller (DF) (1979) test gives the specification as

$$y_t = a_1 y_{t-1} + \varepsilon_t \quad [or\ \Delta y_t = \gamma y_{t-1} + \varepsilon_t] \quad (51)$$

where $a_1 = 1 + \gamma$. The conditions for stationarity require that the null hypothesis of $a_1 = 1$ is rejected in favor of the alternative hypothesis of $-1 < a_1 < 1$. As such, series $\{y_t\}$ decays to the attractor $y^* = 0$.

The problem regarding this test is that the DF test, and even its augmented version, the ADF test, may give a biased estimate in the presence of nonlinearity. Pippenger & Goering (1993) and Balke & Fomby (1997) find that "tests for unit root have low power in the presence of asymmetric adjustment," as cited in Enders (2014, p. 461). As we know, the series of the triangular arbitrage appears to be stationary; the Johansen tests suggest that the spot exchange rate is an unbiased estimator to the BX rate across the period of Regime 2 and Regime 3. The triangular arbitrage series, as the excess values between the log of the BX rate and the log of the FX spot, is expected to have an attractor of zero. However, if the

mean-reverting process is asymmetric, linear models like the ADF test and the Johansen test could fail to detect the attractor due to misspecification.

Hypothesis 4.4 The adjustment of the triangular arbitrage series to the attractor is asymmetric.

4.4.1 Methodology. First, the ADF test is conducted for unit roots; the model with proper length of lags is then selected, removing autocorrelation in disturbances. Rejecting the null indicates series stationarity. The residuals should pass the Ljung-Box test establishing no significant autocorrelation.

For testing nonlinearity, the autocorrelation function (ACF) as used in linear models may be misleading.

The regression error specification test (RESET) posits the null hypothesis of linearity against the general alternative of nonlinearity. From the best-fitting linear model (the ADF model), we extract the residual sequence, denoted $\{\varepsilon_t\}$, and the fitted values, denoted $\{\hat{y}\}$. We then use $H = 4$ to estimate the regression equation

$$\varepsilon_t = \delta z_t + \sum_{h=2}^{H} \alpha_h \hat{y}_t^h \tag{52}$$

where z_t is a vector contains the constant and the regressors used in the linear regression model. If the F-statistic for $\alpha_2 = \cdots = \alpha_H = 0$ exceeds the critical value from the standard F-table, the null hypothesis of linearity is rejected in favor of the nonlinearity alternative (Enders, 2014, p. 415)

The McLeod & Li (1983) test examines the autocorrelation existing in the squared disturbances from a linear regression. Here, the Ljung-Box statistic is used:

$$Q = T(T+2)\sum_{i=1}^{n} \rho_i^2/(T-i) \tag{53}$$

where ρ_i denotes the coefficient of correlation between the squares of the estimated disturbances, $\hat{e}_t^2$ and $\hat{e}_{t-i}^2$.

Enders & Granger (1998) generalized the DF test to a threshold autoregressive (TAR) model given by

$$\begin{gathered}\Delta y_t = I_t\gamma_1(y_{t-1}-\tau) + (1-I_t)\gamma_2(y_{t-1}-\tau) + \varepsilon_t \\ I_t = \begin{cases} 1 & if \;\; y_{t-1} \geq \tau \\ 0 & if \;\; y_{t-1} < \tau \end{cases}\end{gathered} \tag{54}$$

where I_t is the indicator function.

When $y_{t-1} = \tau$, $\Delta y_t = 0$; if $y_{t-1} \geq \tau$, $\Delta y_t = \gamma_1(y_{t-1}-\tau)$; if $y_{t-1} < \tau$, $\Delta y_t = \gamma_2(y_{t-1}-\tau)$. Parameter τ is the attractor since when $y_{t-1} = \tau$, $\Delta y_t = 0$. If $\gamma_1 = \gamma_2 = 0$, the model reduces to a random walk model. If $-2 < \gamma_1 + \gamma_2 < 0$, the process is stationary. Note the DF test is nested in this test when $\gamma_1 = \gamma_2$. Rejecting the null $\gamma_1 = \gamma_2 = 0$ indicates that there is an attractor. Enders (2014) provides the critical values for the F-statistics on $\gamma_1 = \gamma_2 = 0$.

4.4.2 Results. For the series of triangular arbitrage in the Regime 2 period, the best fitting ADF model is represented as

$$\begin{gathered}\Delta r\_ta_t = \underset{(0.04)}{-0.383 r\_ta_{t-1}} - \underset{(0.04)}{0.232\Delta r\_ta_{t-1}} - \underset{(0.03)}{0.106\Delta r\_ta_{t-2}} + \varepsilon_t \\ \text{AIC} = -6601.81 \quad \text{BIC} = -6582.49\end{gathered} \tag{55}$$

The coefficient of r_ta_{t-1} has a t-statistic of -10.78, exceeding the 1% critical value provided by the ADF test; hence, the null hypothesis of a unit-root is rejected. The Q (20)

statistic of the Ljung-Box test is suggestive of no autocorrelation found in 20 lags of the residuals.

The RESET test with $H = 4$ has a p-value of 0.003, rejecting the null hypothesis of linearity at the 1% significance level. The Mcleod-Li test is supportive of nonlinearity: The Q (4) statistic has a value of 28.88, rejecting the null hypothesis of the absence of autocorrelation in the squared residuals.

Chan (1993) proposes an approach to finding a consistent estimate of the threshold called grid searching. The general point is to consider the residual sum of squares (SSR) of a TAR model as a function of the threshold. For any possible values of r_ta_{t-1}, we estimate the TAR model in (54) until we find the one that has the smallest SSR value. To maintain sufficient samples in each subset, we set the trimming value equal to 15%. After trimming, there are 648 potential values of τ to estimate. As shown in Figure 12, when $\tau^* = 0.0001$, the estimated TAR model has the minimum RSS (0.04).

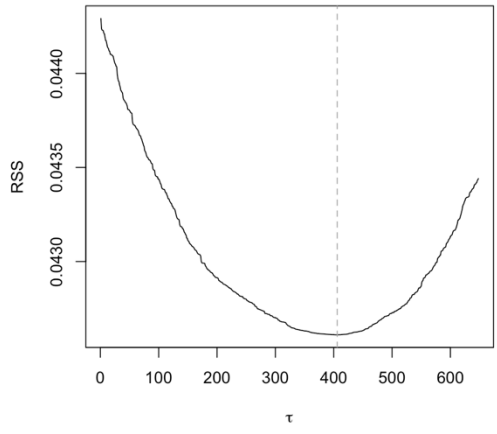

Figure 12. Plot of the RSS of the estimated TAR model on the potential values of τ. The RSS has a minimum value of 0.04 at the 406th point, where $\tau = 0.0001$, as indicated by the grey dashed line.

As $\tau^* = 0.0001$, the estimated TAR model is represented as

$$\Delta r\_ta_t = -0.420 I_t (r\_ta_{t-1} - 0.0001) - 0.338(1 - I_t)(r\_ta_{t-1} - 0.0001) - 0.232\Delta r\_ta_{t-1} - 0.107\Delta r\_ta_{t-2} + \varepsilon_t \quad (56)$$

$$\text{AIC} = -6601.51 \quad \text{BIC} = -6577.36$$

Equation (56) shows that the long-run equilibrium value of 0.0001 is an attractor. When $r\_ta_{t-1} = 0.0001$, Δr_ta_t decays to zero. The adjustment process appears to be asymmetric. If $r\_ta_{t-1} \geq 0.0001$, Δr_ta_t is equal to $-0.420\left(r_{ta_{t-1}} - 0.0001\right) - 0.232\Delta r_{ta_{t-1}} - 0.107\Delta r\_ta_{t-2}$, indicating a faster adjustment process. When $r\_ta_{t-1} < 0.0001$, Δr_ta_t is equal to $-0.338(r\_ta_{t-1} - 0.0001) - 0.232\Delta r\_ta_{t-1} - 0.107\Delta r\_ta_{t-2}$, indicating a relatively slow adjustment process.

The value of the F-statistic for $\gamma_1 = \gamma_2 = 0$ is 59.03; the 1% critical values for T=250 reported by Enders are 8.14 for the TAR model with one lagged change and 8.35 for the model with four lagged changes, respectively. Since the null hypothesis of $\gamma_1 = \gamma_2 = 0$ is rejected at the 1% level, it can be concluded that there is an attractor equal to 0.0001. However, the AIC and BIC information criteria prefer the linear ADF model to the nonlinear TAR model. We then employ the test for the asymmetric adjustment, i.e., $\gamma_1 = \gamma_2$ using the F-statistic. The Dickey-Fuller test is nested in the TAR model; when $\gamma_1 = \gamma_2$, it emerges. The F-statistic for the null hypothesis $\gamma_1 = \gamma_2$ is 1.70, with a p-value of 0.19. We next conduct the F-test for $\gamma_1 = \gamma_2$ on the ADF model with the attractor. The result suggests that the null hypothesis $\gamma_1 = \gamma_2$ cannot be rejected; the F-statistic is 2.13, with a p-value of 0.14. Hence, we can conclude the triangular arbitrage series (or the excess value between the BX and FX spot series) has an attractor equal to 0.0001; the adjustment around it is symmetric.

The TAR model for Regime 3 is estimated as

$$\Delta r\_ta_t = 0.0003 - 0.422 I_t\left(r_{ta_{t-1}} + 0.0017\right)$$

$$- 0.593(1 - I_t)\left(r_{ta_{t-1}} + 0.0017\right) + \sum_{i=1}^{4} \alpha_i \, \Delta r\_ta_{t-i} + \varepsilon_t \qquad (57)$$

$\text{AIC} = -5516.37 \quad \text{BIC} = -5482.26$

For Regime 3, we have 525 observations after omitting 5 outliners. The attractor is equal to -0.0017. The value of the F-statistic for $\gamma_1 = \gamma_2 = 0$ is 42.23, which is significant at the 1% level.

5 Modelling II: In the short run

This chapter investigates the short-run dynamics of the bitcoin exchange rate, the foreign exchange rate, and the triangular arbitrage.

5.1 Vector error correction model

The Johansen test has an equivalent vector error correction model (VECM) representation. Specifically, in presence of co-integration, matrix $\mathbf{\Pi}$ can be decomposed into $\boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\beta}$ is the cointegrating vector and $\boldsymbol{\alpha}$ is the speed-of-adjustment vector.

Like the inference on $\boldsymbol{\beta}$ which is introduced in section 4.2, we can conduct inference on $\boldsymbol{\alpha}$, or on both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$H_3: \boldsymbol{\alpha} = \boldsymbol{A}\boldsymbol{\psi} \tag{58}$$

$$H_4: \boldsymbol{\beta} = \boldsymbol{H}\boldsymbol{\varphi} \ \ and \ \boldsymbol{\alpha} = \boldsymbol{A}\boldsymbol{\psi} \tag{59}$$

The restrictions can be made by constraining on $\boldsymbol{H}$ and $\boldsymbol{A}$.

5.1.1 Results. The VECM representations from the estimated Johansen tests are given by:

(i) Regime 1

$$\Delta bx\_ue_t = -1.106(bx\_ue_{t-1} - 0.258ue_{t-1} - 0.221) + \sum_{i=1}^{2} \gamma_{1i}\Delta bx\_ue_{t-i} + \varepsilon_{1t}$$
$$\Delta ue_t = 0.001(bx\_ue_{t-1} - 0.258ue_{t-1} - 0.221) + \sum_{i=1}^{2} \gamma_{2i}\Delta ue_{t-i} + \varepsilon_{2t} \quad (60)$$

(ii) Regime 2

$$\Delta bx\_ue_t = -0.380(bx\_ue_{t-1} - ue_{t-1} + 0.001) + \sum_{i=1}^{3} \gamma_{1i}\Delta bx\_ue_{t-i} + \varepsilon_{1t}$$
$$\Delta ue_t = 0.005(bx\_ue_{t-1} - ue_{t-1} + 0.001) + \sum_{i=1}^{3} \gamma_{2i}\Delta ue_{t-i} + \varepsilon_{2t} \quad (61)$$

(iii) Regime 3

$$\Delta bx\_ue_t = -1.070(bx\_ue_{t-1} - 1.002ue_{t-1} + 0.001) + \sum_{i=1}^{3} \gamma_{1i}\Delta bx\_ue_{t-i} + \varepsilon_{1t}$$
$$\Delta ue_t = -0.084(bx\_ue_{t-1} - 1.002ue_{t-1} + 0.001) + \sum_{i=1}^{3} \gamma_{2i}\Delta ue_{t-i} + \varepsilon_{2t} \quad (62)$$

Note that $\boldsymbol{\beta}'\boldsymbol{y}_{t-1} = 0$ is the long-run equilibrium relation and $\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$ represents the one-period-ahead deviation from the equilibrium value. Each entry of vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ becomes the speed-of-adjustment parameter multiplying with the historical deviation $\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$.

As suggested by the estimated parameters in (60), for one-unit deviation from the equilibrium relation, the current change of bx_ue_t appears to decrease 110.6% of the deviation, while the current change of ue_t appears to increase 0.1% of the deviation. The speed of adjustment for the BX rate is fairly rapid, even over responding by 10%. This result indicates that the bitcoin exchange rate tends to oscillate widely around the equilibrium value in Regime 1.

The equations in (61) report that the current change bx_ue_t appears to decrease by 38% of the one-period-before deviation, while the current change of ue_t appears to increase by 0.5%. Together, the two current changes eliminate about 38.5% deviation within one day.

Table 13. Inferences on $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ that is estimated from the Johansen tests over Regime 2 and Regime 3.

	$\boldsymbol{H_1}$ $\boldsymbol{\Pi = \alpha\beta'}$	$\boldsymbol{H_{3.1}}$ $\hat{\alpha}_1 = 0$	$\boldsymbol{H_{3.2}}$ $\hat{\alpha}_2 = 0$	$\boldsymbol{H_4}$ $\hat{\beta}_1 = -\hat{\beta}_2$, $\hat{\beta}_0 = 0$, and $\hat{\alpha}_2 = 0$
	Regime 2			
$Q(H_i\|H_1)$		79.44***	0.03	1.56
$\hat{\beta}_0$	-0.000	-0.004	-0.000	0.000
$\hat{\beta}_1$	1.000	1.028	0.999	1.000
$\hat{\beta}_2$	-1.000	-1.000	-1.000	-1.000
$\hat{\alpha}_1$	-0.380	0.000	-0.384	-0.379
$\hat{\alpha}_2$	0.001	0.151	0.000	0.000
	Regime 3			
$Q(H_i\|H_1)$		184.04***	3.97**	22.47***
$\hat{\beta}_0$	0.001	-0.004	0.002	0.000
$\hat{\beta}_1$	1.002	1.025	1.003	1.000
$\hat{\beta}_2$	-1.000	-1.000	-1.000	-1.000
$\hat{\alpha}_1$	-1.070	0.000	-0.992	-0.922
$\hat{\alpha}_2$	-0.084	0.253	0.000	0.000

Note: $Q(H_i|H_1)$ denotes the Q statistic for the likelihood ratio test in (41). The null hypotheses H_i, for $i = 3.1, 3.2, or\ 4$ specified in row 2 of this table, is tested against the alternative hypothesis H_1, the presence of one cointegrating vector. $\widehat{\boldsymbol{\beta}}$ is normalized so that $\hat{\beta}_2 = -1$.
** and *** significant at 5% and 1%, respectively.

In the period of Regime 3, the speed-of-adjustment parameter for the BX rate is about -1.070, indicating the rate of decrease of 107% for one day; the change of ue_t also decreases by 8.4%. Totally, two changes response for 98.5% of the deviations.

Table 13 reports the results of inference regarding $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. For Regime 2, $H_{3.1}: \hat{\alpha}_1 = 0$ is rejected, whereas $H_{3.2}: \hat{\alpha}_2 = 0$ and $H_4: \hat{\alpha}_2 = 0, \hat{\beta}_0 = 0\ and\ \hat{\beta}_1 = -\hat{\beta}_2$ cannot be rejected, so that the FX spot rate is an unbiased estimator to the BX rate; only the BX series responds to the discrepancy, giving a -37.9% response to the one-period-before

equilibrium error. For regime 3, all hypotheses are rejected significantly, hence both bx_ue_t and ue_t respond to the previous deviation, giving a -107% and -8.4% responses, respectively.

Moreover, for Regime 2, only the BX series responds to the discrepancy from the long-run equilibrium relationship, while the FX spot series does not, a situation called weak exogeneity. As such, the spot series is weakly exogenous. Since the FX spot series is a random walk and weakly exogenous, $E(ue_t|\Phi'_{t-1}) = ue_t$. Hence, the conditional expectation of the excess value between bx_ue_t and ue_t equal to zero: $E(bx\_ue_t - ue_t|\Phi'_{t-1}) = ue_t - ue_t = 0$. This suggests that the short-run deviation can be considered a "fair game" regarding the historical and publicly available information set Φ'_{t-1}.

5.2 Impulse response function

The impulse response function (IRF) provides a way to trace the time path of the various shocks on the variables contained in a VAR system. The theory is based on the idea that a vector autoregression can be transformed to a vector moving average (VMA) and that the accumulated effects of two serially uncorrelated error terms can be calculated by this moving average representation. By plotting the IRF, the behavior of the BX and spot series in response to orthogonal shocks is depicted visually. When there is cointegration, the VAR model is misspecified due to not extracting the co-integrating terms from the one-period-lagged autoregressive term. Hence, the appropriate approach is to augment the VAR model with a VECM term.

The estimated Johansen model can be transformed into an VECM model, then represented in a VMA.

5.2.1 Results. The rows in Figure 13 show the IRF of each BX series and the FX spot series plotted with a 10-period horizon. The two shocks, ε_{1t} and ε_{2t}, relate to the two equations in which bx_ue_t and ue_t as dependent variables regress on their previous lags, respectively. Note that the VAR model is transformed from the Johansen's model so that the error-correction terms are also included, which means that it is not a pure unrestricted VAR model. In the left two graphs, the shocks come from ε_{1t}, while in the right two graphs, the shocks come from ε_{2t}. Either ε_{1t} and ε_{2t} is serially uncorrelated; however, the two may be serially correlated, so that their economic meanings are difficult to identify.

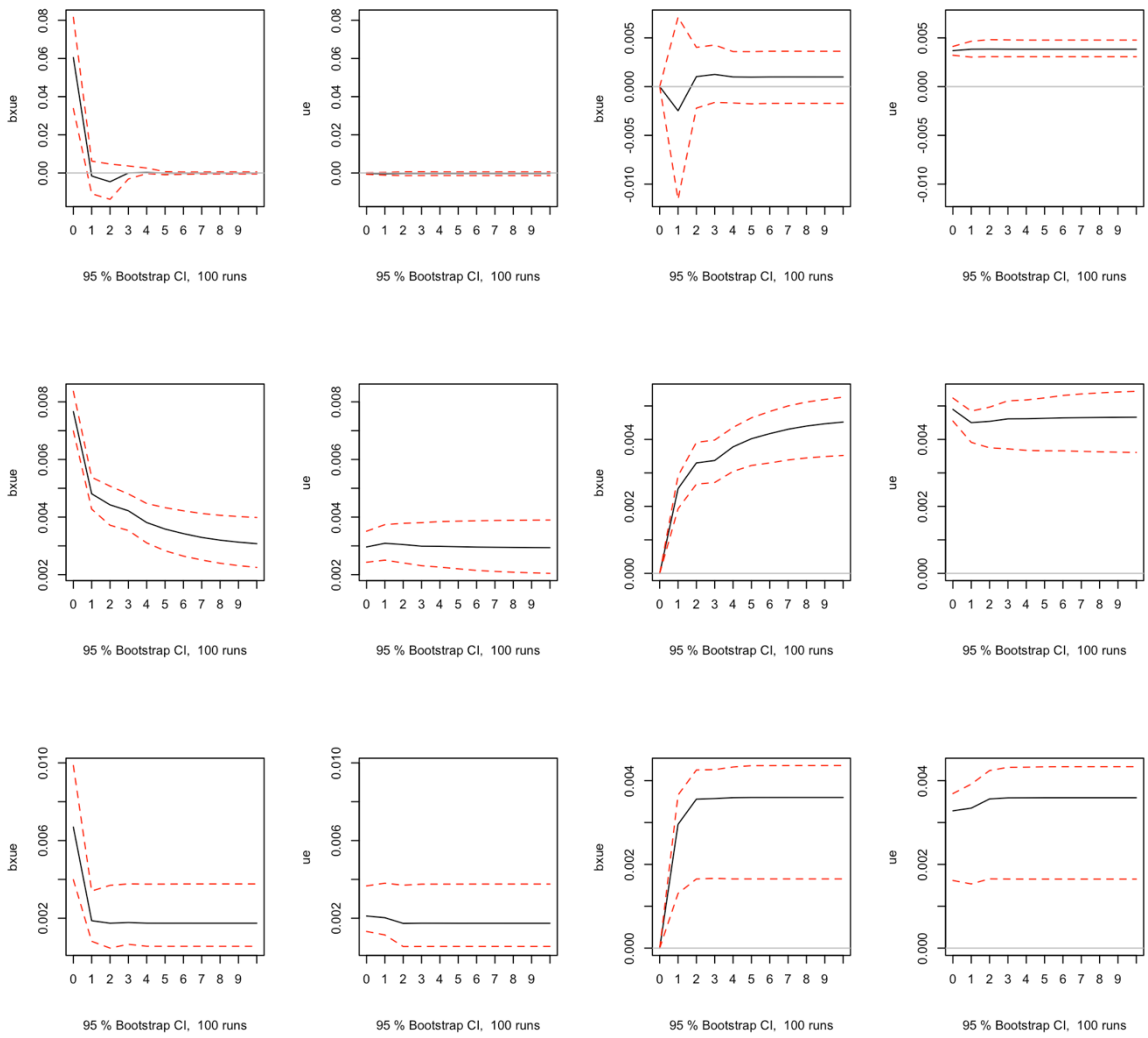

Figure 13. Impulse responses of bx_ue_t and ue_t based on the Johansen tests (Row 1-3 correspond to Regime 1-3, respectively).

In the period of Regime 1, given a one-standard-deviation bitcoin exchange rate shock to ε_{1t}, the BX series increased immediately to above 0.06 at time zero, while the spot series had no response. This result suggests that ε_{1t} and ε_{2t} were not serially correlated at the time, and the sudden appreciation of Euros in bitcoin markets did not exert any influence on the FX market.

Since regardless of whether the BX or spot series in level are unit root processes, all shocks should be permanently cumulated. The FX spot series reflected this feature. However, the BX series did not follow this rule; rather, it adjusted swiftly to eliminate the discrepancy from the long-run equilibrium relationship given by $bx\_ue_t = 0.221 + 0.258ue_t$. As indicated by the value of the speed-of-adjustment coefficient $\hat{\alpha}_1 = -1.070$, the BX series fell more than the quantity of the shock over two periods, causing overreaction. After three periods, it reverted to its before-shock position.

When a one-standard-deviation FX spot shock was given to ε_{2t}, the FX series increased by approximately 0.004 at time zero; it then maintained its position at the new level. The reaction of bx_ue_t lags one period, and nothing happens initially, indicating the lack of correlation of the errors. After fluctuating for three periods, bx_ue_t gradually returned to a new level. Note that the two levels of the BX and FX spot series did not converge, suggesting the change of the FX spot caused the two series were in new equilibrium given by $bx\_ue_t = 0.221 + 0.258ue_t$.

For Regime 2 and 3, the BX series and the FX spot are in a long-run equilibrium relationship of 1:1, so that the two series always came to the same level at the end. The two shocks, ε_{1t} and ε_{2t}, appear to be unidirectionally serially correlated, i.e., given a BX shock, both series increased at time zero, while only the FX spot series responded to the FX shock at time zero. The bitcoin markets seem to be less volatile during the Regime 2 and 3 period, because the one-standard-deviation BX shock caused bx_ue_t to increase by approximately

0.008 and 0.007, respectively. The BX series of Regime 2 presented a fairly slow adjustment process, taking more than eight periods for full convergence. In contrast, the BX series of Regime 3 completed its adjustment in one or two periods.

By observing the behavior of the variables, we are able to establish the following:

(i) The presence of cointegration makes the randomly-walked BX and FX spot series no longer walk randomly; instead, they always approach the level given by their long-term equilibrium relationship.

(ii) In its short-run adjustment, the FX spot series always dominates the level or the equilibrium. Hence, the BX series has to change to compensate for the deviation caused by the shock.

(iii) The absolute value of the speed-of-adjustment coefficient determines the periods required for finishing the error-correcting process. The BX series of the Regime that has the largest speed-of-adjustment coefficient (greater than one in quantity) overreacts to the shock, causing oscillation. The slowest reaction is from the BX series in the period of Regime 2, needing about seven periods. In the period of Regime 3, the BX series completes its error correction process within one period.

5.3 Asymmetric adjustment

The short-run adjustment process described by the VECM and the IRF gives a clear picture of how the BX and FX spot series behave in the presence of deviations. For instance, in Regime 2, the BX series responds by roughly 38% of the one-period-before discrepancy. This assumes that the adjustment is linear or is has the same speed over the period. This is not always the case. From the long view, the speed-of-adjustment at least depends on the

which regime; this result is suggestive of nonlinearity, as the speed-of-adjustment coefficient appears to be "time-varying."

The attractor model introduced in Section 4.4 also provides evidence of a nonlinear adjustment process: The change in the triangular arbitrage series presented different speeds when the one-period-before level series was lower or greater than the threshold value. The presence of asymmetric adjustment makes some linear models, such as the ADF and ARMA models, misspecified.

Next, it is possible to let the adjustment depend on the change in the triangular arbitrage series. As we know, the BX series and the FX spot series are in a 1:1 relationship over the period from Regime 2 to Regime 3; the triangular arbitrage series becomes the equilibrium error. A previous positive change in the error series may exert a different influence on the current change, compared with a negative one. Specifically, a positive equilibrium error stands for the appreciation of Euros in the bitcoin market with respect to the FX market, beneficial to European investors; a negative equilibrium error is suggestive of an appreciation of dollars, beneficial to U.S. investors.

Hypothesis 5.1 The triangular arbitrage series adjusts differently depending on whether it is increasing or decreasing.

5.3.1 Methodology. Caner & Hansen (2001) extend Enders and Granger's (1998) TAR model into the momentum-threshold autoregressive model (M-TAR):

$$\Delta y_t = I_t \gamma_1 (y_{t-1} - \tau) + (1 - I_t) \gamma_2 (y_{t-1} - \tau) + \sum_{i=1}^{p} \beta_i \, \Delta y_{t-i} + \varepsilon_t \qquad (63)$$

$$I_t = \begin{cases} 1 & if\ \Delta y_{t-1} > 0 \\ 0 & if\ \Delta y_{t-1} \leq 0. \end{cases}$$

This specification posits that the series may exhibit more momentum in one direction than the other. The F-statistic for the null hypothesis $\gamma_1 = \gamma_2 = 0$ is called Φ_M.

5.3.2 Results. Under the specification of the M-TAR model in (63), the grid searching approach finds the best fitting $\tau = -0.0005$ where the RSS, AIC and BIC are at their minimum (see Figure 14).

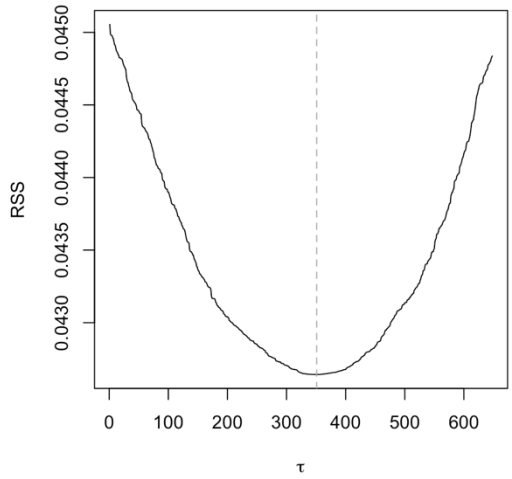

Figure 14. Plot of the RSS for the estimated M-TAR model in (59) on the potential values of τ. The minimum RSS of 0.04 occurred at $\tau = -0.0005$, as indicated by the grey dashed line.

The estimated M-TAR model is represented as

$$\Delta r\_ta_t = -0.389 I_t(r\_ta_{t-1} + 0.0005) - 0.383(1 - I_t)(r\_ta_{t-1} + 0.0005) - 0.229 r\_ta_{t-1} - 0.105 \Delta r\_ta_{t-2} + \varepsilon_t \quad (64)$$

$\text{AIC} = -6600.79 \quad \text{BIC} = -6576.64.$

The previously estimate TAR model for the attractor is given by

$$\Delta r\_ta_t = -0.420 I_t(r\_ta_{t-1} - 0.0001) - 0.338(1 - I_t)(r\_ta_{t-1} - 0.0001)$$
$$- 0.232\Delta r\_ta_{t-1} - 0.107\Delta r\_ta_{t-2} + \varepsilon_t$$

$$\text{AIC} = -6601.51 \quad \text{BIC} = -6577.36$$

Comparing the two results, we find that the M-TAR model does not improve on the TAR model based on the AIC and BIC information criteria. For the M-TAR model, γ_1 and γ_2 have very similar values, -0.389 and -0.383, indicating Δr_ta_t responds equivalently whether the momentum is positive or negative. The F-statistic for the null hypothesis $\gamma_1 = \gamma_2 = 0$ is 58.63, leading to the rejection of the null hypothesis of no attractor. The F-statistic for $\gamma_1 = \gamma_2$ is 0.98, with a p-value of 0.32, indicating that the null hypothesis of an asymmetric adjustment cannot be rejected.

5.4 Nonlinear error correction model

The Granger (1986) representation theory proposes that the Johansen test has the VECM form

$$\Delta bx\_ue_t = -0.396(bx\_ue_{t-1} - ue_{t-1} + 0.002) + \sum_{i=1}^{3} \gamma_{1i} \Delta bx\_ue_{t-i} + \varepsilon_{1t}$$

$$\Delta ue_t = 0.013(bx\_ue_{t-1} - ue_{t-1} + 0.002) + \sum_{i=1}^{3} \gamma_{2i} \Delta ue_{t-i} + \varepsilon_{2t}.$$

After restricting $\beta_0 = 0$ and $\beta_1 = -\beta_2$, we have

$$\Delta bx\_ue_t = -0.391(bx\_ue_{t-1} - ue_{t-1}) + \sum_{i=1}^{3} \gamma_{1i} \Delta bx\_ue_{t-i} + \varepsilon_{1t} \tag{65}$$

$$\Delta ue_t = 0.014(bx\_ue_{t-1} - ue_{t-1}) + \sum_{i=1}^{3} \gamma_{2i} \Delta ue_{t-i} + \varepsilon_{2t} \tag{66}$$

where $bx\_ue_{t-1} - ue_{t-1} = r\_ta_{t-1}$.

Both bx_ue_t and ue_t are $I(1)$ processes, but their linear combination $bx\_ue_{t-1} - ue_{t-1}$ becomes stationary.

Hypothesis 5.2 For any previous deviation from the equilibrium relation presented as the equilibrium errors, the dynamic adjustment could be nonlinear.

5.4.1 Methodology. We can extend the VECM into the M-TAR form:

$$\Delta bx\_ue_t = I_t\gamma_{11}(r\_ta_{t-1} - \tau) + (1 - I_t)\gamma_{12}(r\_ta_{t-1} - \tau) + \sum_{i=1}^{p} \beta_{1i}\,\Delta bx\_ue_{t-i} + \varepsilon_{1t} \tag{67}$$

$$\Delta ue_t = I_t\gamma_{21}(r\_ta_{t-1} - \tau) + (1 - I_t)\gamma_{22}(r\_ta_{t-1} - \tau) + \sum_{i=1}^{p} \beta_{2i}\,\Delta ue_{t-i} + \varepsilon_{2t} \tag{68}$$

$$I_t = \begin{cases} 1 & if\ \ \Delta r\_ta_{t-1} > 0 \\ 0 & if\ \ \Delta r\_ta_{t-1} \le 0. \end{cases} \tag{69}$$

Equations (65) assumes that Δbx_ue_t depends on whether $\Delta r\_ta_{t-1} > 0$ or $\Delta r\_ta_{t-1} \le 0$; Equations (66) assumes that Δue_t depends on whether $\Delta r\_ta_{t-1} > 0$ or $\Delta r\_ta_{t-1} \le 0$. Threshold τ indicates the attractor.

5.4.2 Results. The indicator is given by the TAR model ($\tau = 0.0001$). The resulting nonlinear error-correction model for Regime 2 is represented as

$$\underset{}{\Delta bx\_ue_t} = \underset{(0.05)}{-0.383 I_t(r\_ta_{t-1} + 0.0001)} - \underset{(0.05)}{0.355(1 - I_t)(r\_ta_{t-1} + 0.0001)}$$
$$\underset{(0.04)}{-0.201\Delta bx\_ue_{t-1}} + \underset{(0.05)}{0.149\Delta ue_{t-1}} - \underset{(0.04)}{0.090\Delta bx\_ue_{t-2}} + \underset{(0.05)}{0.124\Delta ue_{t-2}} + \varepsilon_{1t}$$

$$\underset{}{\Delta ue_t} = \underset{(0.04)}{0.008 I_t(r\_ta_{t-1} + 0.0001)} + \underset{(0.04)}{0.011(1 - I_t)(r\_ta_{t-1} + 0.0001)}$$

$$+0.039\Delta bx\_ue_{t-1} - 0.072\Delta ue_{t-1} + 0.015\Delta bx_{ue\,t-2} - 0.015\Delta ue_{t-2} + \varepsilon_{2t}$$
$$(0.03) \qquad (0.04) \qquad (0.03) \qquad (0.04)$$

where the standard errors are in parentheses and two lags of each variable are used in each equation.

Note that all t-statistics in the first equation are significant while no t-statistic is significant in the second equation. These results imply that only the change of the BX series responds to the discrepancy. When the one-period-before excess value between the BX and FX spot series is increasing (i.e. if $\Delta r\_ta_{t-1} > 0$), the BX rate will decrease by 38.3% of the discrepancy between r_ta_{t-1} and the attractor of -0.0001. When the excess value is decreasing, the BX rate will increase by 35.5% of the discrepancy. The adjustment process appears to be asymmetric with respect to whether the excess value is increasing or decreasing.

The linear error-correction model in (65) and (66) suggests that the BX rate always responds to the discrepancy from the long-run equilibrium by 39.1% of the discrepancy.

6 Trading strategies

This chapter introduces two trading strategies based the USD/EUR bitcoin exchange rate and the triangular arbitrage return.

6.1 Trading the bitcoin exchange rate

Nan & Kaizoji (2020) propose an effective arbitrage strategy in bitcoin markets where participants engage in currency trading using the bitcoin exchange rate and hedging the trading risk with the FX futures contract. Their trading strategy is illustrated as follows.

Suppose that an American speculator thinks the Euro has deprecated against the U.S. dollar in the bitcoin market and is betting on the reversion of the situation that may occur in a month. This indicates that the USD/EUR bitcoin exchange rate, denoted $(USD/EUR)_t^{BX}$, is expected to increase in a near future. As a result, the trader places a long position of some Euros at time t, and she may wish to hedge b_t proportion of this fixed cash position in the FX future market, so she goes short b_t on Euros.

The (logarithmic) return on portfolio X comprising of cash and futures at the next period can be represented by:

$$x_{t+1} = r\_bx\_ue_{t+1} - b_t\, r\_fu\_ue_{t+1} \tag{70}$$

where $r\_bx\_ue_{t+1}$ denotes the return on holding a Euro between t and $t+1$; $r\_fu\_ue_{t+1}$ denotes the return on holding the futures position; x_{t+1} denotes the return on holding

portfolio X; b_t is the hedge ratio, defined as the value of futures sales at time t divided by the value of the cash position at time t (Baillie & Myers, 1991).

The naïve hedge ratio is chosen by setting $b_t = 1$. However, this predetermined, fully hedged position is criticized on a problem called over hedging (Cecchetti et al., 1988). The OLS-based hedge ratio is selected by regressing the historical series of return from the asset being hedged on these for the futures to be used. This method minimizes the unconditional variance over the sample period. However, it suffers from three major shortcomings: (i) its objective is to minimize risk, not maximizing the expected utility (Cecchetti et al., 1988); (ii) the joint distribution of prices and futures changes as new data come in (R. F. Engle & Sheppard, 2001); and (iii) the linear regression model faces misspecification when the two regressors are $I(1)$ and cointegrated (Kroner & Sultan, 1993).

6.2 The optimal hedge ratio

The optimal hedge ratio is the slope coefficient, b_t , in (70) that maximizes a speculator's utility. The mean-variance expected utility function, denoted $EU(\cdot)$, is given by

$$EU(x) = E(x) - \gamma Var(x) \tag{71}$$

where $\gamma > 0$ denotes the user's degree of risk aversion and $Var(\cdot)$ denotes the variance operator (Kroner & Sultan, 1993).

Maximize (71), so we have

$$\begin{aligned} \max_b EU(x) = \max_b \{ & E(r\_bx\_ue_t) - b\mathrm{E}(r\_fu\_ue_t) - \gamma [Var(r\_bx\_ue_t) \\ & + b^2 Var(r\_fu\_ue_t) - 2b\, Cov(r\_bx\_ue_t, r\_fu\_ue_t)]\} \end{aligned} \tag{72}$$

where $Cov(\cdot)$ denotes the covariance operator.

Mathematically, the maximum occurs at the point where the first-order derivative of (71) with respect b is equal zero. After solving this difference equation, we have the utility-maximizing hedge ratio given by

$$b' = \frac{2\gamma Cov(r\_bx\_ue_t, r\_fu\_ue_t) - E(r\_fu\_ue_t)}{2\gamma Var(r\_fu\_ue_t)}. \quad (73)$$

The variance of portfolio X returns is expressed as

$$Var(x) = Var(r\_bx\_ue_t) + b^2 Var(r\_fu\_ue_t) - 2bCov(r\_bx\_ue_t, r\_fu\_ue_t). \quad (74)$$

Maximize (74) with respect b, so we have the risk-minimizing hedge ratio expressed as

$$b'' = \frac{Cov(r\_bx\_ue_t, r\_fu\_ue_t)}{Var(r\_fu\_ue_t)}. \quad (75)$$

Note that if the futures rate follows a martingale where $E(\ r\_fu\_ue_t) = 0$, (73) is equal to (75), we get

$$b^* = b' = b'' = \frac{Cov(r\_bx\_ue_t, r\_fu\_ue_t)}{Var(r\_fu\_ue_t)} \quad (76)$$

where b^* denotes the optimal hedge ratio, featuring both minimum risk and maximum mean-variance expected utility (Kroner & Sultan, 1993).

6.3 Measuring the time-dependent variance-covariance matrix

The critical step to calculate the optimal hedge ratio in (76) is to estimate the variance-covariance matrix of the BX rate and the FX futures rate. Using the unconditional variances and covariance implies that the speculator holds the portfolio over the entire sample period and the risks and the correlation coefficient between cash and futures are expected to be all time invariant. Specifically, rewrite (76), we obtain

$$b^* = \frac{Cov(r\_bx\_ue_t, r\_fu\_ue_t)}{Var(r\_fu\_ue_t)} = \frac{Cov(r\_bx\_ue_{\ t}, r\_fu\_ue_{\ t})}{\sigma_{r\_bx}\sigma_{r\_fu}} \frac{\sigma_{rbx}}{\sigma_{rfu}} = \rho_{rbx,rfu} \frac{\sigma_{rbx}}{\sigma_{rfu}} \quad (77)$$

where σ_{rbx} denotes the unconditional standard deviation of the BX return; σ_{rfu} denotes the unconditional standard deviation of the FX futures return; and $\rho_{rbx,rfu}$ denotes the Pearson correlation coefficient between the BX return and the FX futures return.

The time-dependent variances and covariance are more attractive to practitioners than static ones because they can rebalance the portfolio over time rather than holding a fixed position throughout the period (Nan & Kaizoji, 2020). In this case, a bivariate GARCH model serves the purpose of modeling the joint conditional distribution of the returns of the BX and FX futures rates.

There is a variate of multivariate GARCH models. The BEKK model (Engle & Kroner, 1995) is simple to apply but its coefficient matrices are difficult to interpret financially. The GJR model (Glosten et al., 1993) is devised for an asymmetric joint distribution, but because trading on the bitcoin exchange rate can start from any currency and the adjustment of its returns towards the long-run equilibrium appears to be symmetric, asymmetry is not a substantial problem.

The dynamic conditional correlation (DCC-) GARCH model (Engle & Sheppard, 2001; Engle, 2002) contains two components: one for one for the GARCH effect on the conditional covariance matrix and the other for the time-dependent correlations. The DCC-GARCH model is plausible for a large portfolio because the relatively computational ease of the two-step estimation: starting from the univariate GARCH model for each return, then estimating them based on likelihood functions. Although the method is nonlinear, the meaning of the estimated coefficients is straightforward: one set of estimates is for the univariate GARCH environment and the other is for the time-dependent correlations (Engle, 2002).

6.4 VECM plus DCC- GARCH model

Nan and Kaizoji (2019a, 2020) propose to apply the VECM plus DCC-GARCH model to capture the time-dependent covariance matrix regarding the returns of the BX and FX futures rate.

The two variables that we aim to model are return series. As we have known, both the BX rate and the FX futures rate are $I(1)$ and their return series are $I(0)$. It is plausible that first use a VAR model to capture the conditional mean jointly and then pass the residuals from the VAR model to the DCC-GARCH model.

Because the BX rate is found to cointegrated with the FX futures rate (Nan & Kaizoji, 2019b), the unstructured VAR model with the series in first-order differences is misspecified (Engle & Granger, 1987). Therefore, the VAR model needs to incorporate an error correcting term (ECT) to ensure that the long-run equilibrium is maintained in the bivariate system (Kroner and Sultan, 1993), which becomes the bivariate vector error correction model (VECM) (Engle & Granger, 1987).

Usually, the GARCH family works with a hypothetic normal distribution (Bollerslev, 1986; R. F. Engle, 1982). Since the return of the BX rate represent heavy tails, as shown in chapter 3, a Student's t distribution is recommended to incorporate into the DCC-model in order to have better fitting (Nan & Kaizoji, 2020).

Their framework is as follows:

(i) VECM

$$\boldsymbol{r}_t = \boldsymbol{a_0} + \boldsymbol{a}\,\boldsymbol{b}'\,\boldsymbol{y}_{t-1} + \sum_{i=1}^{p} \boldsymbol{c}_i \boldsymbol{r}_{t-i} + \boldsymbol{e}_t \tag{78}$$

where $\boldsymbol{r}_t$ denotes the 2×1 vector of the returns of the BX and the FX futures rates; $\boldsymbol{y}_{t-1}$ denotes the 3×1 vector comprising of a constant, the level BX rate and the level futures rate, one period before; and $\boldsymbol{a}\, \boldsymbol{b}'\, \boldsymbol{y}_{t-1}$ represents the error correction term.

(ii) Probability density assumptions

$$\boldsymbol{e}_t | \Omega_{t-1} \sim Std(0, \boldsymbol{H}_t, \boldsymbol{v}) \tag{79}$$

where Ω_{t-1} denotes the one-period-before information set; $Std(\cdot)$ denotes the standardized student's t density with zero mean, covariance matrix $\boldsymbol{H}_t$ and shape parameter vector $\boldsymbol{v}$.

(iii) DCC-GARCH (Engle, 2002; Engle & Sheppard, 2001)

$$\boldsymbol{H}_t \equiv \boldsymbol{D}_t \mathbf{P}_t \boldsymbol{D}_t \tag{80}$$

$$\boldsymbol{D}_t^2 = diag(\boldsymbol{\omega}) + diag(\boldsymbol{\alpha})\, \boldsymbol{e}_t\, \boldsymbol{e}_t' + diag(\boldsymbol{\beta}) \boldsymbol{D}_{t-1}^2 \tag{81}$$

$$\boldsymbol{\varepsilon}_t = \boldsymbol{D}_t^{-1} \boldsymbol{e}_t \tag{82}$$

$$\boldsymbol{Q}_t = \overline{\boldsymbol{Q}}(\boldsymbol{u}' - \boldsymbol{\Phi} - \boldsymbol{\Psi}) + \boldsymbol{\Phi} \boldsymbol{\varepsilon}_{t-1}\, \boldsymbol{\varepsilon}_{t-1}' + \boldsymbol{\Psi}\, \boldsymbol{Q}_{t-1} \tag{83}$$

$$\mathbf{P}_t = diag\left(\boldsymbol{Q}_t^{1/2}\right)^{-1} \boldsymbol{Q}_t\, diag\left(\boldsymbol{Q}_t^{1/2}\right)^{-1} \tag{84}$$

Conditional covariance matrix $\boldsymbol{H}_t$ in (79) is composed into $\boldsymbol{D}_t \mathbf{P}_t \boldsymbol{D}_t$ where $\boldsymbol{D}_t$ is the diagonal matrix of time-varying standard deviations from the univariate GARCH models, and $\mathbf{P}_t$ is the conditional correlation matrix. The operator $diag(\cdot)$ creates a diagonal matrix from a vector. Equation (81) describes the conditional nature of the volatility. Vector $\boldsymbol{\varepsilon}_t$ in (82) is the standardized innovations. Equation (83) gives the dynamic structure of the conditional correlation matrix using a proxy process $\boldsymbol{Q}_t$, where $\boldsymbol{u}$ is a vector of unities and $\overline{\boldsymbol{Q}}$ is the unconditional correlation matrix of the standardized innovations in (82); this specification assumes that $\boldsymbol{Q}_t$ is integrated and has an exponential smoothing structure (Nan & Kaizoji, 2020). Ding and Engle (2001) shows that restrictions on $(\boldsymbol{u}' - \boldsymbol{\Phi} - \boldsymbol{\Psi})$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ can make $\boldsymbol{Q}_t$ positive semi-definite or positive definite so that the conditional correlation matrix $\mathbf{P}_t$ can be exacted by rescaling $\boldsymbol{Q}_t$ as shown in (84).

The log-likelihood function of a DCC-GARCH model, denoted $LL(\theta, \varphi)$, can be maximized using two steps: first find the estimates of θ in the volatility part of the log-likelihood function, denoted $LL_V(\theta)$, where θ is the parameters in associated with $\boldsymbol{D}_t$ in (81); then, passing $\hat{\theta}$ to $LL_C(\theta, \varphi)$, the correlation part of the log-likelihood function, and maximizing the function with respect to φ, where φ denotes the parameters in (83). The log-likelihood function $LL(\theta, \varphi) = LL_V(\theta) + LL_C(\theta, \varphi)$ can also be estimated using the generalized method of moments (GMM) method (Newey & McFadden, 1994; Engle, 2002).

6.5 Empirical results

The dataset contains the BX rate and the FX futures rate in the Regime 2 period from 2 April 2014 to 9 January 2018 (928 observations). The lag selection is based on the AIC information criterion through the unrestricted VAR model.

Table 14 reports the results. The statistics from the Johansen model with 3 lags show that there exists one cointegrating vector. The long-run equilibrium relationship is given by $bx\_ue_t - 1.004 fu\_ue_t + 0.0067$; the equilibrium error is retained as the error-correction term for the VECM model in (78).

In the VECM model, a_{0B} and a_{0F} are constants; a_{1B} and a_{1F} are the speed-of-adjustment coefficients for the changes of bx_ue_t and fu_ue_t, respectively, and their values imply that the change of bx_ue_t responds to the discrepancy by -28.63% while the change of fu_ue_t responds by 0.34%.

Table 14. Estimation of the bivariate VECM plus DCC-GARCH model

VECM		**Bivariate GARCH** (1, 1)		**Conditional correlation and Information criteria**	
b_0	0.0067	ω_B	0.0000***	φ	0.0971***
			(0.00)		(0.03)
b_1	-1.0040	ω_F	0.0000	ψ	0.8881***
			(0.00)		(0.05)
lags	3	α_B	0.1318***		
			(0.02)		
a_{0B}	-0.0001	α_F	0.0494***	LL	6974.73
			(0.01)		
a_{0F}	-0.0001	β_B	0.7397***	AIC	-14.99
			(0.03)		
a_{1B}	-0.2863	β_F	0.9492***	BIC	-14.85
			(0.01)		
a_{1F}	0.0034	v_B	4.4930***		
			(0.55)		
		v_F	6.0724***		
			(1.12)		

Note: In the VECM model, coefficients b_0 and b_1 specify the long-run equilibrium relation obtained from the Johansen test, specifically, $BX_t + b_0 + b_1 FU_t = e_t$, where e_t denotes the equilibrium error.
The lag length is determined by the Akaike Information Criterion (AIC) through the unrestricted VAR model of the BX and FX futures series in level.
LL denotes Log-Likelihood, and AIC denotes the Akaike Information Criterion for the framework.
*** significant at 1%.

After the VECM filtration of the unconditional mean, autoregression in the first moment, and the error correcting term, the volatility part of the DCC-GARCH model shows the mean-reverting pattern of the volatility of the BX rate returns, with $\alpha_B + \beta_B = 0.8715$, and the integrated pattern of the volatility of the FX futures returns, with $\alpha_F + \beta_F = 0.9986$. For the BX return, $\alpha_B = 0.1318$ and $\beta_B = 0.7397$ suggest that the current conditional variance is highly correlated with the previous conditional variance but is less correlated with the previous filtered squares of the error. For the FX futures return, the conditional variance process presents a fairly persistent pattern. The estimated shape coefficients of Student's t density suggest that $v_B = 4.49$ and $v_F = 6.07$, so the BX returns present heavier tails than the FX futures returns.

The conditional proxy process $\boldsymbol{Q}_t$ appears to be persistent, as well, according to the magnitudes of φ and ψ, with $\varphi + \psi = 0.9852$.

As plotted in Figure 15, the graph of the conditional variance of the BX rate return, $h_{B,t}$, shows mean-reverting and volatility clustering features, while the graph of the conditional variance of the futures return, $h_{F,t}$, appears to show a random walk process. The conditional covariance remains generally positive (above the red dashed line) with some dramatic negative spike.

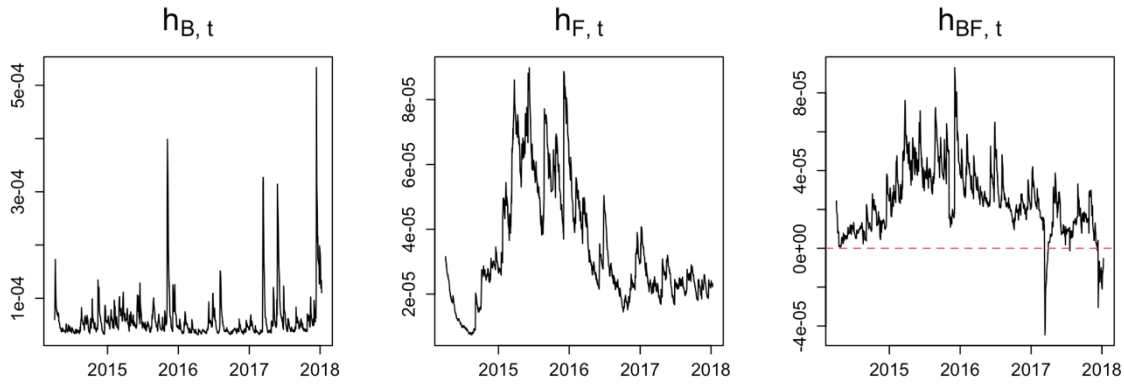

Figure 15. The conditional variance of the return of the bitcoin exchange rate ($h_{B,t}$), the conditional variance of the return of the FX futures ($h_{F,t}$) and the conditional covariance ($h_{BF,t}$).

The plot of b_t^* (see Figure 16) shows a time-varying pattern, moving up and down around the red dashed line, which represents the conventional variance-minimized hedge ratio calculated by the OLS method ($b_c = 0.637$).

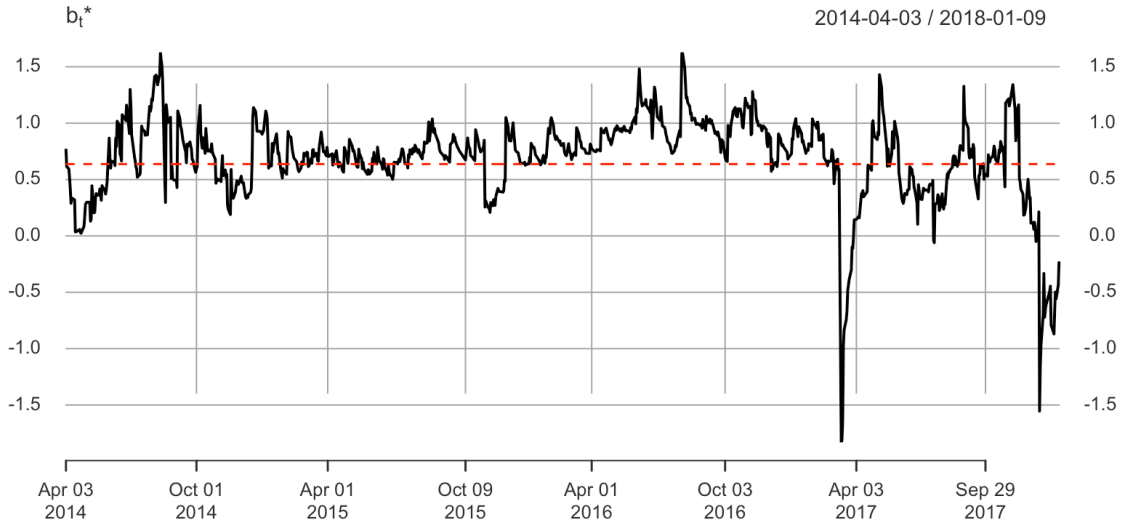

Figure 16. The conditional optimal hedge ratio (b_t^*) in the black line and the conventional hedge ratio given by the OLS method in the red dashed line.

The static hedge ratio given by the OLS method suggests that about 63.7% cash position should be sheltered from currency risk. The optimal number of futures contracts needed to hedge the cash position is calculated as

$$\frac{Units\ of\ the\ cash\ postion \times The\ optimal\ hedge\ ratio}{The\ size\ of\ one\ futures\ contract}. \tag{85}$$

Usually, the value of optimal hedge ratio is between zero and one. However, the resulting optimal hedge ratio ranges from -1.83 to 1.63; a value greater than one implies an extreme volatility; a negative hedge ratio suggests the speculator would instead hedge by shorting futures contracts. Hence, the results show that the FX futures can normally hedge the risk from bitcoin-based currency trading, but in some extreme cases, this strategy is not enough.

6.6 User utility considering the transaction costs

Following the method proposed by Kroner and Sultan (1993), Nan & Kaizoji (2020) construct a conditional user utility using the conditional variances obtained from the VECM + DCC GARCH model and take account of transaction costs. The scenario is that when the utility gained from changes in the variances is enough to offset the transaction cost, the user rebalances; when the gain is less than the cost the user remains in the previous position.

The condition for rebalancing is given by

$$-C-\gamma\left(h_{B,t+1}-2b_{t+1}^{*}\,h_{BF,t+1}+b_{t+1}^{*2}\,h_{ff,t+1}\right) > -\gamma\left(h_{B,t+1}-2b_{t}^{*}\,h_{BF,t+1}+b_{t}^{*2}\,h_{F,t+1}\right) \tag{86}$$

where $h_{B,t+1}-2b_{t+1}^{*}\,h_{BF,t+1}+b_{t+1}^{*2}\,h_{ff,t+1}$ is portfolio's conditional variance at time t+1 and C denotes the percentage return that the user pays as a cost of the transaction.

Using the data in the Regime 2 period, we calculate the mean-variance utility by summing each individual utility (927 observations). The results are reported as follows.

When costs $C=0.0005$ (or 0.5%), a speculator needs to rebalance her position 3 times according the conditional optimal hedge ratio in order to attain her maximum utility that is equal to -0. 1516 over the sample period.

When costs $C=0.0001$ (or 0.1%), the algorithm suggests that the number of rebalances is 9 and the utility is equal to -0.1544. When costs $C=0.00001$ (or 0.05%), the number of rebalances becomes 79 and the utility is -0.1577.

Nan & Kaizoji (2020) made a comparison of the portfolios based on the naïve, conventional and conditional optimal hedge ratios. Their results showed that the conditional optimal hedging portfolio is superior to the other two portfolios in a few aspects: maximum log- likelihood for model estimation, minimum unconditional variance of the portfolio, minimum Value-at-Risk of the portfolio, and maximum mean-variance utility.

6.7 Triangular arbitrage

Nan & Kaizoji (2019a) argue that 'an arbitrage strategy that relies on the bitcoin exchange rate appears rather primitive due to the one-at-a-time nature of the unidirectional exchanges.' For example, after finding that the Euro has depreciated against the U.S. dollar in the bitcoin market, a speculator traded U.S. dollars for Euros using the bitcoin exchange rate. Then, she had to wait for the appreciation of Euros to change her Euros back to dollars in the bitcoin market. As suggested by the speed-of-adjustment coefficient and the IRF, the adjustment process in the Regime 2 period may take 8 or 9 days. During that period, the speculator would face the problem of a limited budget and the risk of holding Euros.

A triangular arbitrage means that instead of waiting for anther arbitrage opportunity, the speculator trades Euros back for U.S. dollars in the FX spot market. This process can be repeated over and over until the discrepancy between the BX rate and the FX spot has been eliminated or the range of deviations falls into the transaction cost band (meaning that the profits from the triangular arbitrage are less than the transaction costs).

Nan & Kaizoji (2019a) propose an ARMA (1, 1) plus DCC-GARCH (1, 1) model to measure the joint conditional distribution of the triangular arbitrage and the return of the FX futures; the empirical and forecast results suggest that the portfolio containing the bitcoin-based triangular arbitrage and its FX futures hedge is superior to the USD/BTC and EUR/BTC assets in terms of risk management.

7 Forecasting

One of the central goals of modeling financial data is forecasting. This chapter illustrates two examples: volatility forecasting with the VECM + DCC-GARCH model and bitcoin exchange rate forecasting using artificial neural networks.

7.1 Volatility forecasting using the VECM + DCC-GARCH model

If there exists time dependency in a time series, that information could be utilized for forecasting. The bitcoin- related series, such as the bitcoin prices, the bitcoin exchange rate, and the bitcoin-based triangular arbitrage, all present long-range serial dependency in terms of autocorrelation. However, as we have observed, either an ARMA model or a GARCH model is not adequate for modeling the bitcoin-related series. Due to the presence of heavy tails in the distributions of the bitcoin-related returns, the Student's t density is needed to address the problem. The results show that the ARMA (p, q) plus GARCH (1, 1) model with a Student's t density appears to be adequate. Hence, we select to forecast volatility in the bitcoin exchange rate.

7.1.1 Methodology. In chapter 6, the bitcoin exchange rate and the FX futures rate are jointly estimated using the VECM plus DCC-GARCH (1, 1) with a student's t density over the Regime 2 period. Based on the estimated model, the one-step-ahead forecast is simply

$$E_t \boldsymbol{D}_{t+1}^2 = diag(\boldsymbol{\omega}) + [diag(\boldsymbol{\alpha}) + + diag(\boldsymbol{\beta})]\boldsymbol{D}_t^2 \tag{87}$$

where E_t denotes the conditional expectation operator.

Then, using the forward iteration, we get the j-step-ahead forecast function as

$$E_t \boldsymbol{D}_{t+j}^2 = \sum_{i=1}^{j-2} diag(\boldsymbol{\omega}) \, [diag(\boldsymbol{\alpha}) + diag(\boldsymbol{\beta})]^i + [diag(\boldsymbol{\alpha}) + diag(\boldsymbol{\beta})]^{j-1} \boldsymbol{D}_{t+1}^2 \tag{88}$$

where $j \geq 2$.

Note that the quality of the forecast declines as j increases. It is common to use rolling one-step-ahead approach to make a forecast, i.e., the length of the sample window is fixed, and as a new data point come into the window, we throw the oldest data point and estimate the model again.

7.1.2 Forecast evaluation. Volatility is latent so that we need find an ex-post proxy for volatility in order to evaluate the forecast (Brownlees et al., 2011). Two classes of the proxies commonly used are the squared returns and the realized returns. Then, volatility forecast comparison can be conducted through loss functions. The mean squared error (MSE) loss is specified as

$$MSE: \quad L\left(\hat{\sigma}_t^2, h_{t|t-1}\right) = \left(\hat{\sigma}_t^2 - h_{t|t-1}\right)^2 \tag{89}$$

where $h_{t|t-1}$ denotes the one-step-ahead conditional variance and $\hat{\sigma}_t^2$ denotes the ex-post proxy.

Patton (2011) proposes the quasi-likelihood (QL) loss which is said to be robust in generating the same ranking of models as long as the proxy is unbiased. The QL loss is specified as

$$QL \;:\; L\left(\hat{\sigma}_t^2, h_{t|t-1}\right) = \frac{\hat{\sigma}_t^2}{h_{t|t-1}} - log\frac{\hat{\sigma}_t^2}{h_{t|t-1}} - 1. \tag{90}$$

7.1.3. Results. Figure 17 plots the forecasted conditional variance of the returns of the BX rate ($h_{B,t}$), the forecasted conditional variance of the returns of the FX futures ($h_{F,t}$), the forecasted conditional covariance ($h_{BF,t}$), and the forecasted conditional correlation (Q_t) over the Regime 3 period (535 observations).

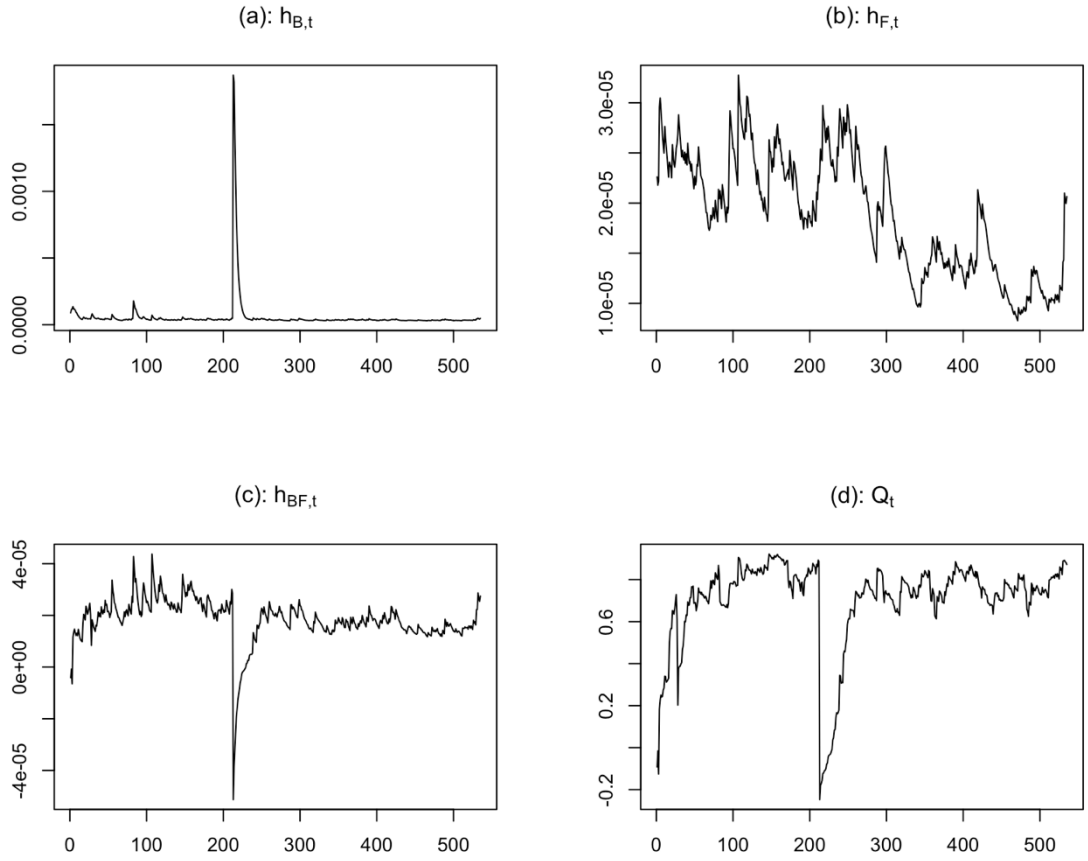

Figure 17. The forecasted conditional variance of the returns of the BX rate ($h_{B,t}$), the forecasted conditional variance of the returns of the FX futures ($h_{F,t}$), the forecasted conditional covariance ($h_{BF,t}$), and the forecasted conditional correlation (Q_t) over the Regime 3 period.

As what we reported in Chapter 6, volatility in the bitcoin exchange rate presents clustering while volatility in the FX futures rate behaves like a random walk process. The mean of the MSE losses and the mean of the QL losses for the forecasted BX rate variance are 7.21×10^{-7} and 2.38, respectively. For forecasted the FX future rate variance, there is only the mean of MSE losses equal to 76.87×10^{-10}, but the mean of QL losses does not exist because some values of the logs become infinity.

The conditional standard deviation series is plotted with the series of the absolute return, the 150-day moving average standard deviation, and the unconditional standard deviation of Regime 2 (see Figure 18).

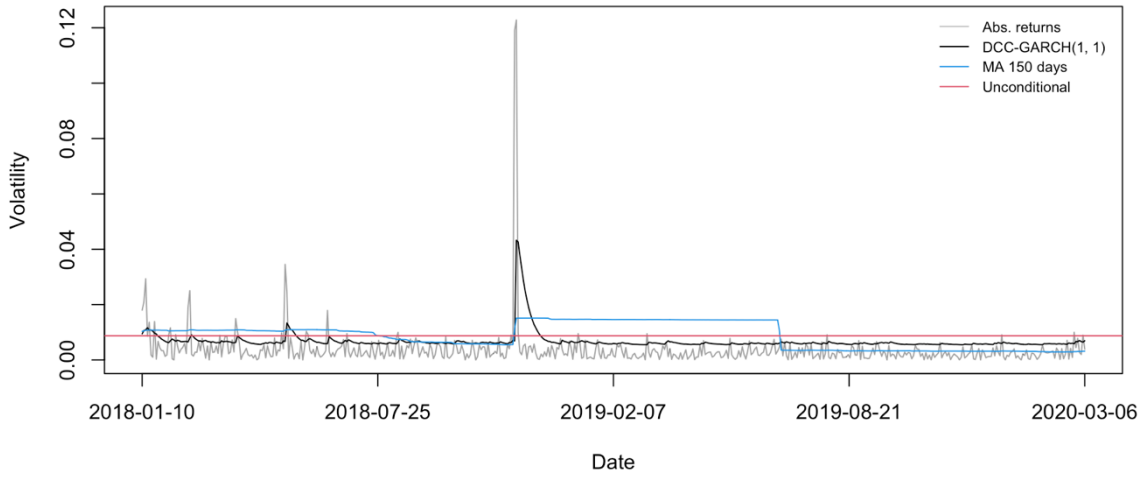

Figure 18. The absolute returns, the forecasted conditional standard deviation, the 150-day moving average and the unconditional standard deviation of Regime 2.

7.2 Forecasting the bitcoin exchange rate using neural networks

Forecasts are typically made on a covariance stationary series such as a return series. The reason for this is that a stationary series presents an equilibrium-reverting feature—given a shock, the effect of the shock is always decaying to zero as time goes on, unless there is a new shock. A linear model assumes that the speed of the decay is constant; a nonlinear model posits that the speed is in a known function form. Hence, forecasting is about finding an optimal decay function that gives minimum errors between the forecasted and the observed values.

There is a trick that makes the price "predictable." For example, assume we have observed a logarithmic price sequence, $y_0, y_1, \ldots, y_t$, denoted $\{y_t\}$, in which the return series is a pure AR (1) process with a zero mean; that is, there are no MA terms and there is no volatility clustering. If we let $a_1 = 0.7$, the AR (1) process is expressed as

$$\Delta y_t = 0.7\Delta y_{t-1} + \varepsilon_t$$

where ε_t is a white-noise process. The one-period-ahead conditional forecast is simply

$$E_t\Delta y_{t+1} = 0.7\Delta y_t \tag{91}$$

Rewrite equation (91), and we get

$$E_t y_{t+1} = y_t + 0.7y_t - 0.7y_{t-1} = 1.7y_t - 0.7y_{t-1}.$$

This strategy works since we know that Δy_t is decaying at a constant speed of 0.7 and our knowledge of the standard deviation of ε_t assures us that the estimate is within a certain confidence interval, say, 95%. This is the fundamental of forecasting in econometric time-series analysis.

The price itself, however, is difficult to forecast. We use the description of a random process offered by Endres (2014, p. 184) to illustrate why this is the case. The efficient market hypothesis posits that the daily changes in the price are completely random. As such, the current price is expressed as

$$y_t = y_{t-1} + \varepsilon_t. \tag{92}$$

Given an initial condition y_0, the general solution to equation (92) is

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i, \tag{93}$$

which means that the current price is the initial price plus the sum of all random errors from time 1 to time t; the effects of stochastic shocks never decay but are accumulated permanently. Taking the expectation of (93), we obtain $Ey_t = y_0$, so that the value of y_0 is an unbiased estimator of all future values of y_1 through y_t. The variance of y_t is such that $Var(y_t) = Var(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) = t\sigma^2$; thus, the value of the variance is a function of time t. As time t goes to infinity, the variance of y_t approaches infinity. Thus, $Ey_{t+1} = y_t$ is the best forecast strategy for a random process, i.e., y_t is an unbiased estimator of the expected value of y_{t+1} and the variance of y_{t+1} is σ^2. This should be the unbeatable common-sense baseline to econometric forecast.

With the help of new technology such as machine learning and the development of computational power, the common-sense baseline can be substantially improved upon. We illustrate how neuron networks can achieve this. There are several categories of machine learning algorithms such as the artificial neural network (ANN), the support vector machine (SVM), decision trees, random forests, gradient boosting machines, etc. In this thesis, we use a "deep" neural network, as it is "feature-engineering"-free and is the approach least dependent on human help to learn features from the data.

A neural network is like a black box with several layers inside (the input layer, the hidden layer(s) and the output layer); given an input, it will produce an output. In each hidden layer, there are several units, called neurons. Each neuron computes $\boldsymbol{w} \cdot \boldsymbol{y} + \boldsymbol{b}$, where $\boldsymbol{w}$ denotes the weight vector, $\boldsymbol{y}$ is the input vector, and $\boldsymbol{b}$ denotes the bias vector. Vector $\boldsymbol{w}$ and $\boldsymbol{b}$ are called hyperparameters. Several such jointing layers constitute a neural network. Neural networks use these hyperparameters to memorize the representation mapping the input and output values.

A neural network can learn from the difference between its output value and the target to find an optimal representation of the input value in terms of the output value, a process called supervised learning. Given a loss function, a neural network tries to minimize the losses between its output and the target during each iteration and adjusts its hyperparameter to achieve the minimization through the back-propagation method. Through successive iterations, the network finally learns how to mathematically represent the target in terms of its hyperparameter. In theory, a neural network with one hidden layer can learn any forms of functions, including probability distributions; a two-hidden-layer neural network can learn conditional probability densities (Husmeier, 1998). If there are many layers, the network is called a deep learning neural network. The reason for incorporating many layers in a neural network is to "achieve great power and flexibility by representing the world as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones" (Goodfellow et al., 2016).

The goal of a learning algorithm is to minimize the loss function. A promising way for achieving this is to follow a direction opposite to the gradient of the loss function. To find the optimal solution, we can compute the gradient of the loss for all data regarding all the parameters of the network; however, this approach is computationally expensive, as, for

some networks, there are millions of parameters (Hegde & Usmani, 2016). Instead, we usually compute the gradient of the loss functions of a subset of the dataset and repeat the procedure over and over until all data have been used up. This iterative process is called an epoch; the subset of the data is called a mini-batch (Hegde & Usmani, 2016). To make the subset more representative of the entire dataset, we sample each subset with shuffling.

After training the network, we face a critical barrier in the neural network field, called overfitting. This problem concerns trading off between the representation of the network and the generalization of the network. The network is well-fitted to the training dataset, learning some patterns as well as noise, but what is the performance of the network when we apply it to a new dataset—the validating dataset—in the process of generalization? Validation is used to address the problem of overfitting. Finally, the neural network is applied to the testing dataset to evaluate its accuracy for forecasting.

This section describes how a neural network trained by a 5-minute dataset can outperform the common-sense baseline formed from the random walk model in one-day-ahead bitcoin exchange rate forecasts.

7.2.1 Methodology. The dataset consists of five time series with a 5-minute basis: the bitcoin exchange rate, the FX spot rate, the return of the bitcoin price of USD, the return of the bitcoin price of EUR, and triangular arbitrage. There are 364,936 observations for each series. The first 200,000 observations of the five series are used for training the network, the next 100,000 observations are retained for validation, and the last 64,936 observations are used for testing. The experiments contain two classes of neural networks: a densely connected network (a feedforward network) and a gated recurrent unit (GRU) network (a recurrent network).

We let the network look back at the data points for five days, meaning 1,440 observations, set the value of the step equal to twelve so that the network samples 120 points out of the 1,440 observations, and let the network forecast the one-day-after bitcoin exchange rate (the next 288th data point). Before feeding the dataset into the network, we rescale each series according its mean and standard deviation.

The size of the mini-batch is 128. For training, we use 20 epochs and 500 steps per epoch; for validation, we set 770 steps. The rmsprop optimizer is used for the stochastic gradient descent algorithm. The loss function is the mean absolute error (MAE) metric.

The neural networks are built using the R version Keras, an open-source neural network library written in Python. The code is modified according to the book by Chollet and Allaire (2018).

The common-sense baseline is constructed by calculating the errors between the current bitcoin exchange rate and the next day bitcoin exchange rate (the next 288th data point) on the validation dataset (the middle 200,000 observations); as noted, the loss function is the MAE metric.

7.2.2 Results. The common-sense baseline method yields an MAE of 0.0925, which means that if we use the one-day-before bitcoin exchange rate to forecast the current rate, we produce a mean absolute error of 0.0925. Since the input bitcoin exchange rate is normalized using the series mean of 1.1545 and standard deviation of 0.0916, the MAE value without normalization is equal to 0.0925×0.0916 = 0.0085, meaning an MAE of 85 basis points.

Model 1is a densely connected model with three layers, as shown in Table 15. The first layer takes in the input data on shape (120, 5) and flattens them into an array of shape

600. The second layer takes in the array and represents it in a 32-dimensional hypothetical space. The third layer gives the forecasted bitcoin exchange rate. The network has 19,265 trainable parameters.

Table 15. Model 1: the densely connected neural network

Layer type	Output shape	Number of parameters
Flatten	600	0
Dense	32	19,232
Dense	1	33

Note: The total number of the parameters is 19,265.

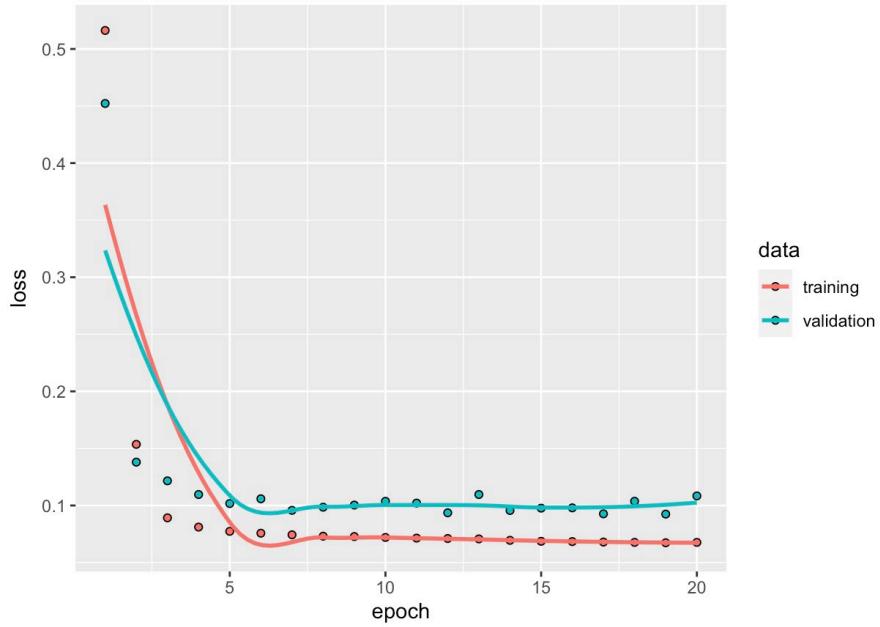

Figure 19. Training and validation losses on the bitcoin exchange rate with the densely connected neural network.

Figure 19 shows the results after training the network for 20 epochs. The red points in the figure are the training losses; the green points are the validation losses. The smoothed points in each series are shown in the colored lines. The training process yields an MAE of 0.5162 at the first epoch, far above the baseline of 0.0925; however, after three epochs, the

loss decreases to 0.0892. The losses continue to decrease as more training epochs are conducted. The validation losses initially decrease, but after five epochs, the losses maintain at a level of approximately 1.0. These results show that it is difficult for a densely connected network to out-perform the common-sense baseline even though the network itself is powerful.

Using a neural network for forecasting involves a trade-off between representation and generalization. This two-layer network (not including the flatten layer) is sufficiently complicated to represent 600 values in a 32-dimensional hypothetical space, but the features that the network learns cannot be generalized to the new data. A possible explanation for this is that the historical "prices" did not provide any useful information for forecasting the future event, as the efficient market hypothesis conjectures. Note that the information (or the features) refers to the values of price and its geometric distribution because no sequential features can be learned by this type of network.

Model 2 is a gated recurrent unit (GRU) network with two layers, as shown in Table 16. The GRU network uses the same sample principle as the long short-term memory (LSTM) network, which recurs the output of the network into its input so that it can learn sequential features, and introduces modules so that the network can have a long memory. However, the GRU network is cheaper to run and has acceptable representational power (Chung et al., 2014).

Table 16. Model 2: the gated recurrent unit neural network

Layer type	**Output shape**	**Number of parameters**
GUR	32	3,648
Dense	1	33

Note: The total number of the parameters is 3,681.

The losses are plotted in Figure 20 and show that the GRU network is fairly powerful for sequence modeling, i.e., the first epoch yields a training MAE of 0.0771 and a validation MAE of 0.0845, out-performing the common-sense baseline. Note that the two plotted lines do not intersect and that after the second epoch, the validation losses start to increase, indicating that the model is overfitting. The minimum validation loss (0.0819) occurs at the second epoch.

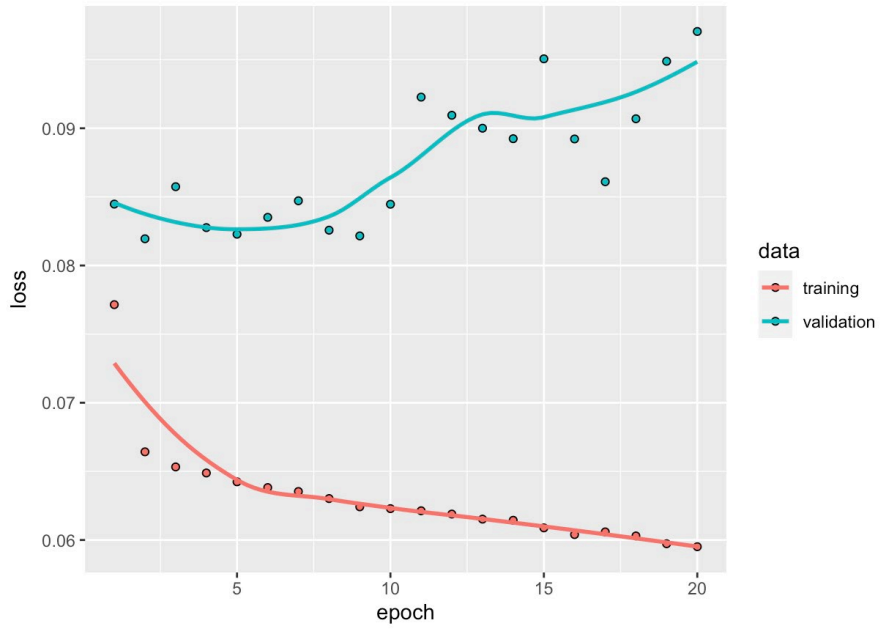

Figure 20. Training and validation losses on the bitcoin exchange rate with the GRU neural network.

Model 3 is a GRU network with two dropout masks. The first dropout mask is set before the recurrent layer with the dropout rate equal to 3%. For the recurrent dropout mask, we use a constant mask, as suggested by Gal (2016), with the dropout rate equal to 7%. As plotted in Figure 21, the line of the training losses intersects the validation loss line at the third epoch; the minimum loss (0.0811) occurs at the sixth epoch. The translated MAE score is equal to 0.0811×0.0916 = 0.0074, or 74 basis points. This is a significant improvement in forecasting accuracy, decreasing the MAE score by (0.0925-0.0811)/0.0925 = 12.32%.

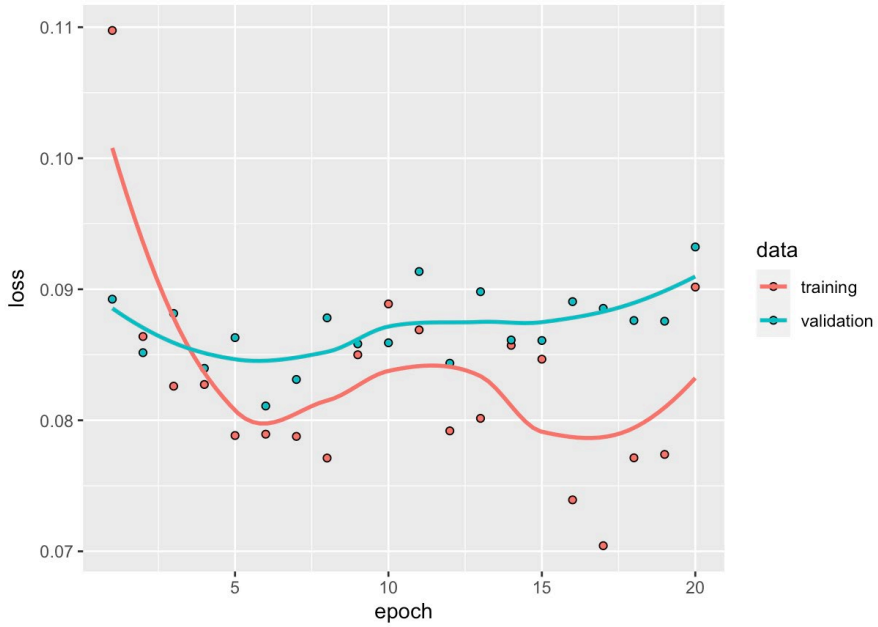

Figure 21. Training and validation losses on the bitcoin exchange rate with the GRU neural network using dropout.

Finally, we apply Model 3, trained for 7 epochs, to the test dataset. Model 3 yields an MAE score of 0.0436, compared to the common-sense baseline MAE of 0.0446.

8 Discussions and Conclusion

This thesis investigates currency trading in bitcoin markets. Specifically, the USD/EUR pair is chosen because of its popularity with traders. The USD/EUR price in bitcoin markets, which is of central interest for both academic researchers and practitioners, is expressed as the USD/EUR bitcoin exchange rate. Based on the observable USD/EUR bitcoin exchange rate series, a variety of quantitative approaches to characterizing, modeling, and forecasting are applied; the bitcoin exchange rate is modeled or compared with other assets such as the bitcoin prices of USD and EUR, the FX spot, futures, forward rates, and the one-month interest rate. The goal of this extensive effort is to gain a global view of bitcoin-based currency trading though a number of different models.

Before concluding, two issues related to the bitcoin exchange rate warrant discussion: the bid-ask spread and the confirmation time.

8.1 The bid-ask spread

The USD/EUR bitcoin exchange rate is constructed using the bitcoin price indices, which represent the mid-point price. Wide bid-ask spreads in bitcoin prices may mean that no trading can occur at the suggested bitcoin exchange rate. The data provided from data.bitcoinity.org suggest that the daily spreads are quite different across markets.

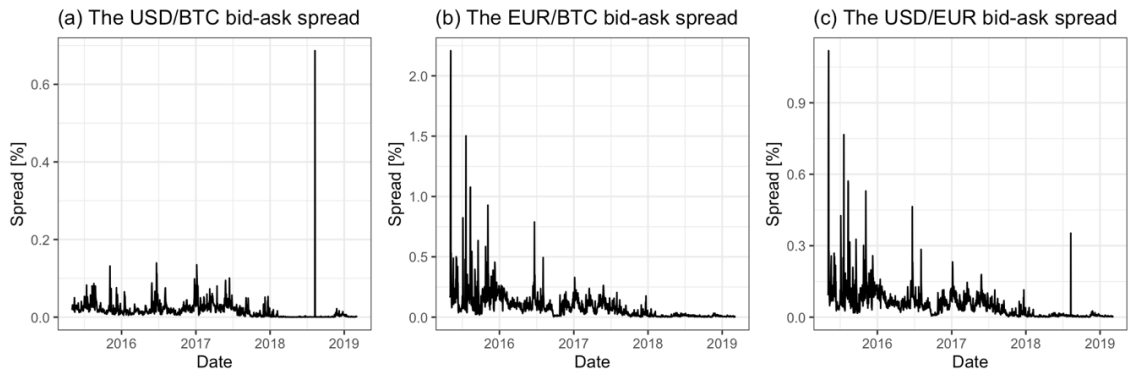

Figure 22. The USD/BTC bid-ask spread, the EUR/BTC bid-ask spread, and bid-ask spread of the USD/EUR bitcoin exchange rate (Coinbase).

Coinbase, which ranks first in terms of market share, has the smallest spread in these markets. Panels (a) and (b) in Figure 22 plot the series of the daily bid-ask spreads of the USD/BTC and the EUR/BTC over the period from 2 May 2015 to 6 March 2020. Though volatile, the spreads are mostly less than 1%. The bid-ask spread of the USD/EUR bitcoin exchange is estimated using (1/2 USD/BTC spread) + (1/2 EUR/BTC spread) (see panel (c) of Figure 22). This time-varying spread appears not to affect the bitcoin exchange rate, which uses the mid-price as an unbiased estimator of the USD/EUR price in the bitcoin market. The reason for this is that if the distribution of the spread is symmetric, the median is equal to the mean, so that the bitcoin exchange rate is unbiased to the mean. As such, the spread can be viewed as a random cost factor added in the bitcoin exchange rate. If the distribution of the spread is asymmetric but not persistent, the spread is still treated as the transaction cost. Only if the distribution is persistently asymmetric towards one side is the bitcoin exchange rate biased. However, such persistence may be thought of as a risk premium, e.g., a liquidity discount for accepting bitcoins (Dong & Dong, 2015).

8.2 The confirmation time of bitcoin transferring

The confirmation time of bitcoin transfer causes problems for the bitcoin exchange rate, as delay will introduce into bitcoin-based currency trading the risk of holding bitcoins. As depicted in Figure 1 in Section 2.1, the maximum average confirmation time exceeded 300 minutes and the median confirmation time varied over a range of approximately 10 minutes. Such long delays in bitcoin transfer may make currency trading using the bitcoin exchange rate catastrophic.

Consider a case in which the deviations between the bitcoin exchange rate and the FX spot rate occurred but no traders are willing to arbitrage because of the risk associated with transfer delay. Then, the deviations suggested by triangular arbitrage should meander anywhere and should not show a reversion to equilibrium in a wide cost-band. The adjustment process can only happen when the profit of arbitrage exceeds the risk of holding the bitcoin during the delay. However, this situation did not occur in our analysis; instead, triangular arbitrage presented mean-reversion over the Regime 2-3 period. This is quite like statistical inference: We have a null hypothesis positing the absence of convergence between the bitcoin exchange rate and the FX spot rate, but convergence that has very little chance of occurring randomly if the null hypothesis is true is shown to occur, leading to a rejection of the "no convergence" null hypothesis. Hence, we conjecture that a long confirmation time for bitcoin transfer does not discourage arbitrage by speculators. One explanation for this is that there is no need for real bitcoin transfers when bitcoins are traded; rather, the transaction occurs in the brokerage account.

Next, we investigate the effect of a time delay using 30-minute data constructed from the 5-minute dataset. Assume that a trader changes one Euro into bitcoins at time 1, denoted $-eb_1$ and then trades the bitcoins for U.S. dollars at time 2 (a half-hour later), denoted ub_2.

Hence, the 30-minute-delayed bitcoin-based USD/EUR price quoted at time 2 is expressed as $ub_2 - eb_1$. Generally, the 30-minute-delayed, bitcoin-based USD/EUR price series is given by $bx\_ue\_d30_t = ub_t - eb_{t-1}$. This series can be compared to the BX series and the other series. Let us define

$$Return\ 1: \quad bx\_ue_t - bx\_ue_{t-1} \tag{94}$$

$$Return\ 2: \quad bx\_ue\_d30_t - bx\_ue_{t-1} \tag{95}$$

$$Return\ 3: \quad bx\_ue\_d30_t - bx\_ue_t \tag{96}$$

$$Return\ 4: \quad bx\_ue\_d30_t - bx\_ue\_d30_{t-1} \tag{97}$$

$$Return\ 5: \quad bx\_ue\_d30_t - bx\_ue\_d30_{t-2} \tag{98}$$

$$Return\ 6: \quad bx\_ue\_d30_t - bx\_ue\_d30_{t-3} \tag{99}$$

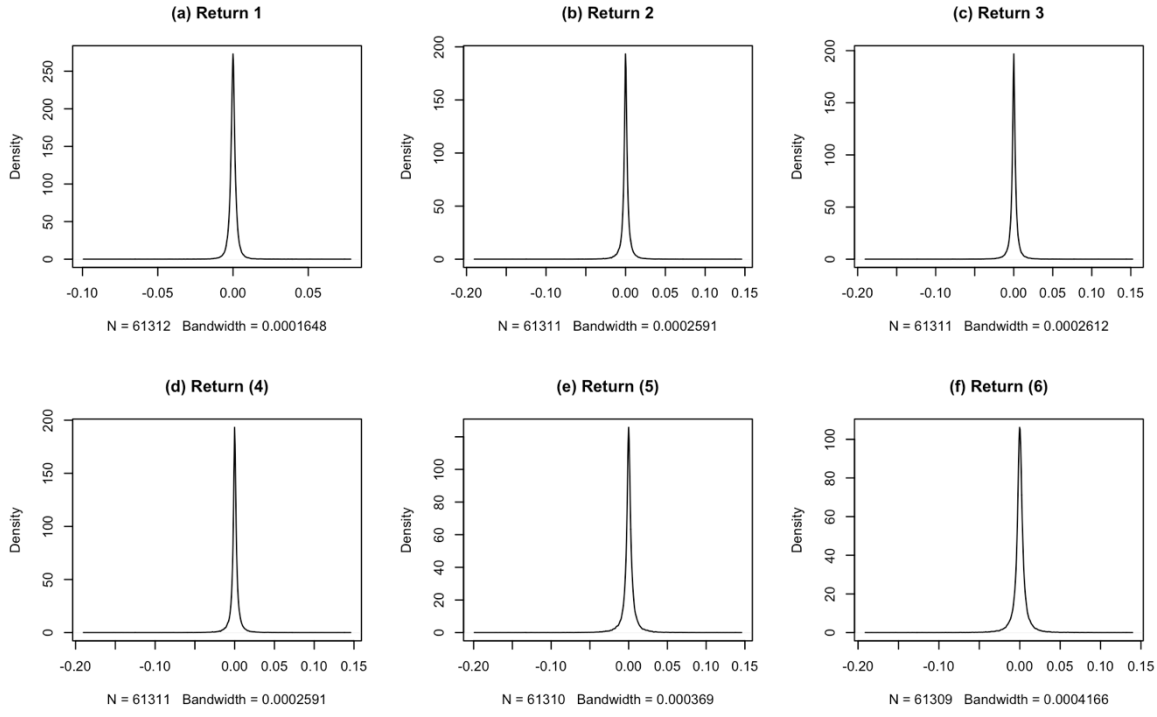

Figure 23. Plots of the series of six returns as defined in equations (94) through (99).

Figure 23 plots the series of the six defined returns. Panel (a) shows the density of the BX rate returns. Return 2 is the logarithmic excess value between the 30-minute-delayed

return and the BX rate quoted before the delay (panel (b)). Return 3 is the logarithmic excess value between the 30-minute-delayed return and the BX rate quoted after the delay (panel (c)). Both logarithmic excess values present a zero mean, indicating that the delay does not affect the mean but does affect the variance. As a result, the BX rate is an unbiased estimator to the delayed prices in the bitcoin markets.

The second row of Figure 23 shows the delayed return density with increasing delay times—30 minutes, 60 minutes, and 90 minutes. Here again, the delays have no effect on the mean.

8.3 Conclusion

Currency trading in bitcoin markets can be performed through two bitcoin transactions. The USD/EUR bitcoin exchange rate defined in Nan and Kaizoji (2017, 2019b) appears to be the best estimator of all possible bitcoin-based USD/EUR rates that considers time delay. Confirmation time does not affect the mean USD/EUR rate but will incorporate more risks. The bid-ask spread issue rate becomes trivial if we think of the spread as a transaction cost.

From their plots, we find that the USD/EUR bitcoin exchange series intertwines with the FX spot series over time, regardless of the wide variations in the two bitcoin price indices. The sample summary statistics are very similar between the bitcoin exchange rate and the FX spot, especially the standard deviations. In the linear specification, $bx\_ue_t = \beta_0 + \beta_1 ue_t + \varepsilon_t$, the OLS estimator β_1 and the correlation coefficient $\rho_{bx,ue}$ for bx_ue_t and ue_t are represented as $\beta_1 = \rho_{bx,ue}(\sigma_{ue}/\sigma_{bx})$; given $\beta_1 = 1$ and $\sigma_{ue} = \sigma_{bx}$, $\rho_{bx,ue}$ must be equal to 1, indicating a perfect linear relationship in the samples of the bitcoin exchange rate and the FX spot rate.

The return of the bitcoin exchange rate has a zero mean, a moderate standard deviation equal to 0.027, a positive skewness of 2.873, and substantial excess kurtosis. The zero mean indicates no risk premium, which distinguishes a martingale from a submartingale. The moderate standard deviation indicates a middle-level risk, coinciding with the results from the 2.5% and 1% VaR. That is, the risk exposure of the bitcoin exchange rate to losses is between that of the bitcoin investment and forex trading. The positive skewness implies heavy asymmetric tails. The leptokurtosis is suggestive of thick tails and a non-normal distribution.

Impressively, both bitcoin price returns have a positive mean of 0.003, indicating a submartingale. Across the entire sample period, this positive mean brings bitcoin traders an average profit of 0.003 × 1600 = 480%. The triangular arbitrage series is the return on the investment of $EUR \rightarrow BTC \rightarrow USD \rightarrow EUR$. The negative mean of -0.001 indicates that the average loss is equal to 0.001 × 1600 = 160%. More generally, we treat the non-zero means as risk premiums—holding bitcoins results in a positive premium, while holding Euros in the triangular arbitrage results in a negative premium.

Our time series analysis gives the stylized features as follows: The bitcoin exchange rate appears to be a random walk process similar to other financial assets, indicating weak form market efficiency. However, the bitcoin-based series returns all represent long memory autocorrelation. This long-range dependence means an ARMA model needing longer AR and MA terms but the result is still model inadequacy, even considering the GARCH effects. Unless the student's t distribution is considered in the framework, the serial correlation problem becomes acceptable to a certain level. The triangular arbitrage series, which assumes the unbiasedness of the FX spot towards the bitcoin exchange rate and the absence of the risk premium, measures how far the bitcoin exchange rate deviates from the FX spot rate. This series does not have a meaningful AR (1) interpretation over the full sample period.

The TAR model using time as the threshold suggests that it is plausible to split the triangular arbitrage into three regimes:

Regime 1: from 11 September 2013 to 1 April 2014, when the bitcoin exchange rate was widely oscillating.

Regime 2: from 2 April 2014 to 9 January 2018, when the bitcoin exchange rate presented a gradually reverting pattern.

Regime 3: from 10 January to 6 March 2020, when the bitcoin exchange rate tended to stay close to the FX spot rate.

This separation suggests structural changes that occurred in the bitcoin markets where the bitcoin price formation gradually took more account of the information from the FX market. It also brings simplicity to the modeling; for example, the triangular arbitrage series is adequately modeled by an AR (3) model.

In the long run, the bitcoin exchange rate was found to cointegrate with the FX spot and futures, respectively. Significantly, the long-run equilibrium relationship between the bitcoin exchange rate and the FX spot rate is suggestive of semi-strong market efficiency over the Regime 2-3 period, and the law of one price appeared to hold during the same period. For Regime 1, the equilibrium relation explains why the two "prices" diverged simultaneously in the bitcoin and FX markets. The CIP relation appeared to hold in a particular context. The nonlinear ADF model succeeds in identifying the attractor existing in the triangular arbitrage series for Regime 2 and Regime 3. The asymmetric adjustment coefficients were estimated based on the attractor, though the statistics appeared not to be statistically significant.

The VECM models depict different short-run dynamics across the regimes. For Regime 1, the bitcoin exchange rate presented independence from the FX spot rate, and over-adjusting occurred from time to time. Over the Regime 2 period, since the FX rate appeared

to be exogenous, the bitcoin exchange rate tended to adjust to the FX spot rate by 40% of the discrepancy in one day. In the period of Regime 3, both rates responded to the discrepancy, and the adjustment was expected to finish within a day. The momentum TAR model and the nonlinear ECM model gave a detailed description of the asymmetric adjustment process. In the case of the bitcoin exchange rate, the adjustment appeared to be symmetric.

It is plausible to use the FX futures to hedge currency risks when traders arbitrage on the bitcoin exchange rate or conduct triangular arbitrage. The proposed bivariate DCC-GARCH models facilitate these trading strategies by providing a time-varying conditional covariance matrix to serve as the basis for the conditional optimal hedge ratio computation. The time-varying conditional correlation coefficient is obtained as a byproduct of the model. When transaction costs are considered, the proposed algorithm that maximizes the user's mean-variance utility gives the optimal number needed to rebalance the portfolio.

Finally, it is essential to investigate the forecast issue. The DCC-GARCH model was used for conditional volatility forecasting using the rolling one-step-ahead approach. In this case, forecasting the bitcoin exchange rate is worth challenging. Impressively, the resulting GRU neural network outperformed the common-sense baseline, the random walk model, on one-day-ahead forecasting using the 5-minute dataset.

Although this thesis employed a variety of models and techniques to characterize, model, and forecast USD/EUR trading in bitcoin markets, it falls short of providing the complete picture. To fully understand the roles that the bitcoin markets play for currency trading, investigations into other currencies are needed (Nan, 2020).

Reference

Akram, Q. F., Rime, D., & Sarno, L. (2008). Arbitrage in the foreign exchange market: Turning on the microscope. *Journal of International Economics*, *76*(2), 237–253.

Ali, R., Barrdear, J., Clews, R., & Southgate, J. (2014). *The Economics of Digital Currencies* (SSRN Scholarly Paper ID 2499418). Social Science Research Network. https://papers.ssrn.com/abstract=2499418

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, 821–856.

Andrews, D. W., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, 1383–1414.

Angelidis, T., Benos, A., & Degiannakis, S. (2004). The use of GARCH models in VaR estimation. *Statistical Methodology*, *1*(1–2), 105–128.

Bai, J., & Perron, P. (2003). Critical values for multiple structural change tests. *The Econometrics Journal*, *6*(1), 72–78.

Baillie, R. T., & Myers, R. J. (1991). Bivariate GARCH estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics*, *6*(2), 109–124.

Balke, N. S., & Fomby, T. B. (1997). Threshold cointegration. *International Economic Review*, 627–645.

Bjerg, O. (2016). How is Bitcoin Money? *Theory, Culture & Society*, *33*(1), 53–72. https://doi.org/10.1177/0263276415619015

Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, *29*(2), 213–38.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Bouri, E., Molnár, P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? *Finance Research Letters*, *20*, 192–198.

Brezo, F., & Bringas, P. G. (2012). *Issues and risks associated with cryptocurrencies such as Bitcoin*.

Briere, M., Oosterlinck, K., & Szafarz, A. (2013). *Virtual Currency, Tangible Return: Portfolio Diversification with Bitcoins* (SSRN Scholarly Paper ID 2324780). Social Science Research Network. https://papers.ssrn.com/abstract=2324780

Brownlees, C. T., Engle, R. F., & Kelly, B. T. (2011). A practical guide to volatility forecasting through calm and storm. *Available at SSRN 1502915*.

Buchholz, M., Delaney, J., Warren, J., & Parker, J. (2012). Bits and Bets, Information, Price Volatility, and Demand for Bitcoin. *Economics*, *312*.

Buchko, S. (2017, December 12). *How Long Do Bitcoin Transactions Take?* CoinCentral. https://coincentral.com/how-long-do-bitcoin-transfers-take/

Caner, M., & Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica*, *69*(6), 1555–1596.

Cecchetti, S. G., Cumby, R. E., & Figlewski, S. (1988). Estimation of the optimal futures hedge. *The Review of Economics and Statistics*, 623–630.

Chan, K.-S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, *21*(1), 520–533.

Cheah, E.-T., & Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, *130*, 32–36.

Chollet, F., & Allaire, J. J. (2018). *Deep Learning mit R und Keras: Das Praxis-Handbuch von den Entwicklern von Keras und RStudio*. MITP-Verlags GmbH & Co. KG.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv Preprint ArXiv:1412.3555*.

Ciaian, P., Rajcaniova, M., & Kancs, d'Artis. (2016). The economics of BitCoin price formation. *Applied Economics*, *48*(19), 1799–1815.

Clinton, K. (1988). Transactions costs and covered interest arbitrage: Theory and evidence. *Journal of Political Economy*, *96*(2), 358–370.

Commodity Futures Trading Commission. (2017). *A CFTC Primer on Virtual Currencies*. Commodity Futures Trading Commission. https://www.cftc.gov/sites/default/files/idc/groups/public/documents/file/labcftc_primercurrencies100417.pdff

Corbet, S., Lucey, B., Peat, M., & Vigne, S. (2018). Bitcoin Futures—What use are they? *Economics Letters*, *172*, 23–27.

Crowder, W. J. (1995). Covered interest parity and international capital market efficiency. *International Review of Economics & Finance*, *4*(2), 115–132.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, *74*(366a), 427–431.

Ding, Z., & Engle, R. F. (2001). *Large scale conditional covariance matrix modeling, estimation and testing*.

Dong, H., & Dong, W. (2015). *Bitcoin: Exchange rate parity, risk premium, and arbitrage stickiness*.

Easley, D., O'Hara, M., & Basu, S. (2019). From mining to markets: The evolution of bitcoin transaction fees. *Journal of Financial Economics*, *134*(1), 91–109.

Elliot, B. E., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests of the unit root hypothesis. *Econometrica*, *64*(8), 13–36.

Enders, W. (2014). *Applied econometric time series* (fourth edition). John Wiley & Son.

Enders, W., & Granger, C. W. J. (1998). Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. *Journal of Business & Economic Statistics*, *16*(3), 304–311.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987–1007.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica: Journal of the Econometric Society*, 251–276.

Engle, R. F., & Kroner, K. F. (1995). Multivariate Simultaneous Generalized Arch. *Econometric Theory*, *11*(1), 122–150.

Engle, R. F., & Sheppard, K. (2001). *Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH*. National Bureau of Economic Research.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Favre, L., & Galeano, J.-A. (2002). Mean-modified value-at-risk optimization with hedge funds. *The Journal of Alternative Investments*, *5*(2), 21–25.

Gal, Y. (2016). Uncertainty in deep learning. *University of Cambridge*, *1*(3).

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M. C., & Siering, M. (2014). *Bitcoin-asset or currency? Revealing users' hidden intentions*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2425247

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.

Granger, Clive W. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, *48*(3), 213–228.

Granger, Clive WJ, & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, *2*(2), 111–120.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*(3), 393–408.

Gün, L. (2014). A New Form of Currency: Description and Economic Principle. *Journal of Scientific Research and Reports, ISSN*, 2320–0227.

Hakkio, C. S., & Rush, M. (1989). Market efficiency and cointegration: An application to the sterling and deutschemark exchange markets. *Journal of International Money and Finance*, *8*(1), 75–88.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, *93*(2), 345–368.

Hegde, V., & Usmani, S. (2016). Parallel and distributed deep learning. In *Tech. Report, Stanford University*.

Hong, K. (2017). Bitcoin as an alternative investment vehicle. *Information Technology and Management*, *18*(4), 265–275.

Husmeier, D. (1998). *Modelling conditional probability densities with neural networks* [PhD Thesis]. Citeseer.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2), 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica: Journal of the Econometric Society*, 1551–1580.

Kang, M. (2019). Currency Market Efficiency Revisited: Evidence from Korea. *International Journal of Financial Studies*, *7*(3), 52.

Kelleher, J. P. (2020, 20). *Why Do Bitcoins Have Value?* Investopedia. https://www.investopedia.com/ask/answers/100314/why-do-bitcoins-have-value.asp

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, *3*, 3415.

Kroner, K. F., & Sultan, J. (1993). Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis*, *28*(4), 535–551.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1–3), 159–178.

Li, Z., Dong, H., Huang, Z., & Failler, P. (2018). Asymmetric effects on risks of Virtual Financial Assets (VFAs) in different regimes: A Case of Bitcoin. *Quantitative Finance and Economics*, *2*(4), 860–883.

Luther, W. J. (2016). Bitcoin and the Future of Digital Payments. *The Independent Review*, *20*(3), 397.

McLeod, A. I., & Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, *4*(4), 269–273.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Bitcoin. Org,[Cit. 2014-11-13]. Dostupnỳ z: Https://Bitcoin. Org/Bitcoin. Pdf*.

Nan, Z., & Kaizoji, T. (2017). Market Efficiency of the Bitcoin Exchange Rate: Evidence from Co-Integration Tests. *Available at SSRN: Https://Ssrn.Com/Abstract=3179981*.

Nan, Z., & Kaizoji, T. (2019a). Bitcoin-based triangular arbitrage with the Euro/US dollar as a foreign futures hedge: modeling with a bivariate GARCH model. *Quantitative Finance and Economics*, *3*(2), 347.

Nan, Z., & Kaizoji, T. (2019b). Market efficiency of the bitcoin exchange rate: Weak and semi-strong form tests with the spot, futures and forward foreign exchange rates. *International Review of Financial Analysis*, *64*, 273-281.

Nan, Z., & Kaizoji, T. (2020). The Optimal Foreign Exchange Futures Hedge on the Bitcoin Exchange Rate: An Application to the US Dollar and the Euro. In *Advanced Studies of Financial Technologies and Cryptocurrency Markets* (pp. 163–181). Springer.

Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconmic time series. *Journal of Monetary Economics*, *10*(2), 139–162.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, *4*, 2111–2245.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Pichl, Lukáš, & Kaizoji, T. (2017). Volatility analysis of bitcoin. *Quantitative Finance and Economics*, *1*, 474–485.

Pichl, L., Nan, Z., & Kaizoji, T. (2020). Time series analysis of ether cryptocurrency prices: Efficiency, predictability, and arbitrage on exchange rates. In *Advanced studies of financial technologies and cryptocurrency markets* (pp. 183-196), Springer, Singapore.

Pippenger, M. K., & Goering, G. E. (1993). A note on the empirical power of unit root tests under threshold processes. *Oxford Bulletin of Economics and Statistics*, *55*(4), 473–481.

van Wijk, D. (2013). What can be expected from the BitCoin. *Erasmus Universiteit Rotterdam*. https://thesis.eur.nl/pub/14100/Final-version-Thesis-Dennis-van-Wijk.pdf

Wewege, L., & Thomsett, M. C. (2019). *The Digital Banking Revolution: How Fintech Companies are Transforming the Retail Banking Industry Through Disruptive Financial Innovation*. Walter de Gruyter GmbH & Co KG.

Yermack, D. (2013). *Is Bitcoin a real currency? An economic appraisal*. National Bureau of Economic Research. http://www.nber.org/papers/w19747