Competition and Product Quality: Fake Trading on Crypto Exchanges

Dan Amiram* Evgeny Lyandres† Daniel Rabetti‡

January 28, 2021

Abstract

How competition affects product quality and how product quality choices impact firms' operating performance are open empirical questions. We use a setting that is especially suitable to answering these questions: We examine the effects of competition among cryptographic exchanges on their trading volume inflation (fake trading) and the impact of volume inflation on exchanges' operating performance. We develop statistical measures to detect fake trading, which we validate in several ways and use in analyzing determinants and consequences of trading volume inflation. Various static and dynamic competition measures are positively associated with measures of fake trading at both the exchange and exchange-currency pair levels. Exchanges that inflate trading volume succeed in misleading investors in the short run but are punished in the long run, consistent with the tradeoff between short-lived increases in rents and future losses due to damaged reputation.

*Coller School of Management, Tel Aviv University, danamiram@tauex.tau.ac.il.
†Coller School of Management, Tel Aviv University, lyandres@tauex.tau.ac.il.
‡Coller School of Management, Tel Aviv University, rabetti@mail.tau.ac.il.
We are grateful to Will Cong, Sanjiv Das, Gustavo Schwenkler, Tong Zhou, and seminar participants at Santa Clara University, Sun Yat-sen University, and University of Wisconsin at Milwaukee for valuable comments and suggestions.

1 Introduction

The choice of product quality has long been recognized in the industrial organization literature as an important facet of operating strategy of firms producing "experience goods", i.e. products whose quality is not observed until after the purchase decision (e.g., Klein and Leffler (1981), Shapiro (1983), Allen (1984), and Cooper and Ross (1984)).[1] A firm's decision whether to attempt to mislead its customers about the quality of its products involves a tradeoff between higher short-term profits and lower future profits due to damaged reputation. Product market competition can be important in shaping this tradeoff. On the one hand, building and maintaining reputation is more valuable the larger the expected future rents relative to current rents. The larger the (expected) competition, which reduces future rents, the lower the incentive to maintain reputation (e.g., Dana and Fong (2011)). On the other hand, the strength of the reputation effect may increase in competition, as more intense competition expands consumers' outside options (e.g., Horner (2002)).

In this paper, we examine empirically the relation between competition and firms' choices of their products' quality, as well as implications of these choices for firms' future performance. We focus on an industry that, as we argue, is well suited to an examination of this relation – cryptographic exchanges whose role is to facilitate trades in pairs of assets at least one of which is a crypto currency. Crypto exchange industry is especially suitable for an examination of determinants and consequences of product quality choices in a competitive setting for several reasons.

First, trading volume on exchanges is positively related to anticipated product quality, as trading volume is generally perceived to be associated with market liquidity. Exchanges have access to an easily implementable technology for misleading consumers regarding the quality of their product: artificial inflation of trading volume by means of fake (wash) trading that is not immediately observed by outsiders (e.g., Cumming et al. (2011)).[2] Our examination of fake trading on crypto exchanges builds on the pioneering work of Cong et al. (2020), who develop several measures of fake trading on crypto exchanges, some of which we adopt and expand on in our study.

[1] See also models of product quality and reputation in particular industries, such as Chemmanur and Fulghieri (1994) in the context of investment banks, and Bolton et al. (2012) and Bar-Isaac and Shapiro (2013) in the context of credit rating agencies.

[2] A crypto exchange can create fake volume in two main ways. First, it can execute back-and-forth transactions among its own wallets. Second, it can crease perverse incentives for investors to trade with themselves, e.g., by eliminating trading commissions for "top-tier" traders.

The reason trading volume is an important signal to market participants of an exchange's quality is that most aggregators of data on cryptocurrencies and crypto exchanges, such as www.CoinMarketCap.com, explicitly rank exchanges on trading volume. In the absence of credible signals of exchanges' quality, many investors rely on these rankings in their choice of trading venue.[3] Consistent with this conjecture, Gervais et al. (2001) find, in the context of traditional exchanges and equities, that unusually high trading volume is followed by abnormally positive returns, suggesting that investors value increased liquidity. Importantly, while wash trading is forbidden in most asset markets,[4] the legality of wash trading in crypto exchanges is a grey area, due to insufficient regulation of crypto markets in most jurisdictions and to the cross-border nature of most crypto exchanges. Crucially, in our setting, it is the crypto exchanges that may have the incentives to inflate volume, whereas in traditional markets it is usually investors who wash trades, e.g., for tax purposes (e.g., Grinblatt and Keloharju (2004)).[5]

Second, there are two crucial elements underlying the relation between product quality choice and reputation: the quality of a firm's product has to be unobservable ex-ante (as is the case of "experience goods"), and it has to be decipherable ex-post, at least partially. Both these elements are present in the crypto exchange setting: the quality of the product (i.e. the real depth of order books and expected slippage) can only be estimated gradually over time.

Third, the crypto exchange industry enables a more nuanced analysis of the relation between competition and product quality. Although young, the industry is very dynamic, with numerous entries and several exits since the establishment of the first crypto exchanges in 2010.[6] Competition among crypto exchanges has many layers. On the one hand, exchanges are heterogenous both operationally – in terms of the set of currency pairs listed on them, and geographically – catering to partially overlapping sets of clients. On the other hand, competition among multiple trading platforms within a given currency pair is largely competition in homogenous goods.

[3]Recently, in response to widespread concerns of trading volume inflation, www.CoinMarketCap.com and other aggregators began to rank exchanges based on additional criteria that are harder if not impossible to manipulate, such as web traffic to an exchange's website, the number of currencies listed and currency pairs traded on the exchange, and the exchange's age.

[4]For example, in the United States, wash trading was made illegal after the passage of U.S. Securities and Exchange Act in 1934 and the Commodity Exchange Act in 1936.

[5]The most recent large-scale case of trading volume inflation occurred on the Canadian exchange CoinSquare, which inflated volume on its platform by 5.5 billion $U.S.

[6]Bitcoinmarket.com went live in March, 2010, followed in July 2010 by Mt. Gox. Both exchanges are now defunct. According to www.CoinMarketCap.com, there are 382 crypto exchanges as of November 2020. Exchanges also participate in mining activities as well as, more recently, underwriting-type activities, facilitating simultaneous issuance and listing of crypto tokens.

We begin by providing a general description of the empirical regularities and evolution of the crypto exchange market. The main data source that allows us to undertake our investigation is www.Kaiko.com, which contains transaction-level data on over 100 largest cryptographic exchanges, for most of them since their inception. After removing exchanges that do not allow trading in any of the most important crypto currencies – Bitcoin (BTC), Ether (ETH), and Tether (USDT) – we are left with 41 exchanges, which are responsible for over 90% of the overall reported crypto trading volume on average.

In line with the exponential growth of the crypto market, whose capitalization increased by more than 2,000 percent between 2013 and 2019, the volume of trading on crypto exchanges, the numbers of crypto currencies and of crypto currency pairs traded on them have also grown tremendously in recent years. For example, the number of distinct currency pairs that can exchanged directly, such as ETH-BTC (Ether to Bitcoin) – a variable that is much harder to manipulate than trading volume – grew from a single pair in 2013 to over 1,700 pairs in the third quarter of 2019.

The development of the crypto market saw a large increase in the competitiveness of the crypto exchange industry: During our sample period, the number of exchanges in our data set increased from one in 2013 to 37 in 2019, and the Herfindahl indices of concentration of the crypto exchange market, measured using the number of traded currency pairs, reported volume of trading, and reported number of trades, decreased from one to roughly 0.1.

In addition to statistics at the exchange level, we present more granular statistics: at the level of traded currency pairs and at the exchange-currency pair level. For example, the mean and median Herfindahl index of volume of trading in a given currency pair across exchanges is roughly 0.25, suggesting strong competition among exchanges on average, whereas the mean and median Herfindahl index of an exchange across currency pairs traded on it is approximately 0.5, suggesting substantial concentration of a typical exchange in a single currency pair.

We proceed by developing measures of product quality, which, in our setting, is inversely related to estimated fake trading. As fake trading is unobservable and not directly detectable without information on addresses of accounts (crypto wallets) that performed the trade, we develop statistical measures of fake trading at the most granular level possible. i.e. at the exchange-currency pair-month level. We estimate all the measures discussed below using both the trading volume and the number of trades series.

Our first measure, also used in Amiram et al. (2015) in the context of detection of errors in financial statements, in Michalski and Stoltz (2013) in the context of detection of errors in macroeconomic data, and, most relevant for our study, in Cong et al. (2020) for detection of wash trading on crypto exchanges, is deviations of the frequencies of first digits of either the trading volume or the number of trades within short time intervals from Benford's Law (Benford (1938)). This law describes expected frequencies of first digit equaling one through nine for data sets obtained by drawing observations from random samples of varying magnitudes. Naturally observed series tend to follow Benford's Law (e.g., Pimbley (2014)). Larger departures from Benford's Law may indicate higher likelihood of data manipulation. Aloosh and Li (2020), who, along with Gandal et al. (2018), provide direct evidence of wash trading by using an internal book of individual trader level records from Mt. Gox exchange, show that Benford's Law is useful in detecting fake trading.

Our second measure of fake trading, which is novel to the literature, is based on the deviation of the distribution of trading volume/number of trades within short intervals from log-normal distribution. Typical trading volume series tend to be distributed log-normally (e.g., Richardson et al. (1986) and Ajinkya and Jain (1989)). Deviations from log-normality can be a sign of data manipulation.

Our third measure of fake trading, which is also novel to the literature, is based on a machine-learning non-parametric algorithm called EDM (E-Divisive with medians), which identifies structural breaks in data series (e.g., James and Matteson (2015) and James et al. (2016)). The larger the number of identified structural breaks in the trading volume series or the number of trades series, the higher the chance of date manipulation e.g., by means of trading volume inflation. This measure, while well known in the computer science literature, is new not only to the crypto exchanges literature but to finance literature in general to the best of our knowledge.

Since the measures discussed above aim to identify particular deviations from normal trading patterns and none of them is likely to be capable of identifying the majority of such deviations, we aggregate these measures by computing their principal components. We use three types of principal components in our empirical analysis: those based on the three measures estimated using trading volume data, those based on the three measures estimated using the number of trades data, and those based on all six measures. While we focus on principal-component-based measures in the analysis, the inferences are largely unchanged when using individual fake trading measures.[7]

[7] While we have data on numerous exchange characteristics potentially associated with fake trading, such as exchange's web popularity, age, location, and regulatory environment, we do not use these characteristics in measuring fake trading, as they cannot be computed at the most granular, exchange-currency pair-month level. We use these characteristics for purposes of external validation of our estimates of fake trading.

Prior to employing these measures in the identification of determinants and consequences of fake trading, we validate the measures in several ways. First, we show that our estimates of fake trading tend to be the lowest for exchanges with the highest web popularity, for the oldest exchanges, and for the minority of exchanges that are regulated to some degree. On the contrary, exchanges with reported cases of manipulation tend to have some of the highest fake trading measures. Second, we find that fake trading measures are higher for exchanges with relatively low levels of self-imposed regulation and compliance, and those with relative low levels of transparency. Third, we estimate other, more intuitive measures of fake trading, which are based on contemporaneous correlations between either the number of trades or trading volume in a given currency pair on a given exchange with the same variable aggregated for that currency pair on all exchanges. The idea is that fake trading, which is unlikely correlated across exchanges, depresses these correlations. We find that our statistical measures of fake trading are significantly associated with alternative, more intuitive, correlation-based measures. Fourth, we use a pseudo-natural experiment – the ban on operation of crypto exchanges in China in 2017, which led to their exodus to other locations, mostly to Hong Kong and the rest of Pacific Asia. We perform a difference-in-differences analysis of changes in fake trading measures on exchanges subject to the ban and find that exchanges that moved out of China following the ban into jurisdictions with higher regulation/compliance standards reduced the extent of their fake trading substantially.

Having validated our measures of trading volume inflation, we use them to examine the relation between static and dynamic aspects of exchanges' competitive environment and the extent of fake trading on their platforms. We report several findings with respect to static competition. At the exchange level, older and larger exchanges tend to exhibit lower fake trading measures, consistent with these exchanges having larger incentives to maintain reputation. At the same time, exchanges that are more diversified across currency pairs tend to have higher estimated fake trading measures, consistent with the result in Dana and Fong (2011) that multimarket competition mitigates reputational damages of opportunistic strategies.

At the exchange-currency pair level, we find that exchanges do not fake trading equally in all currency pairs. Pairs especially prone to volume manipulation are those that are being traded on many exchanges and those in which the concentration of trading across exchanges is low, i.e. in cases in which competition among exchanges is strong. For example, a one-standard-deviation increase in the trading-volume-based Herfindahl index of a currency pair across exchanges is associated with a 0.1-0.2 standard-deviation reduction in fake trading measures in that currency pair. An additional interesting finding is that there is less fake trading in pairs involving tokens issued in an initial coin offering (ICO). A possible reason is exchanges' incentive structure – exchanges are less likely to derive significant ongoing profits from commissions on trading of crypto tokens issued in ICOs relative to commissions on trading crypto coins.

We then examine how dynamic competition environment, i.e. changes in exchanges' competitive landscape, influences their strategies. In particular, we analyze how entry and exit of an exchange's competitors in a given currency pair impact the extent of fake trading of the focal exchange. In doing so, we define three types of competitors: general – any exchange that commences/ends trading in the relevant currency pair; geographical – a subset of competitors operating in the focal exchange's geographical area; and operational – an exchange that has the largest overlap with the focal exchange in the set of currency pairs traded on both. We find that entry (exit) by competitors increases (decreases) fake trading estimates at the exchange-currency pair level, especially when the initial extent of competition that the focal exchange is subject to in a given currency pair is large. Consistent with our findings regarding the effects of static competition on fake trading, this result demonstrates that reputational concerns are decreasing in the extent of competition, in line with Dana and Fong (2011).

The last part of our analysis concerns short-run and longer-run effects of fake trading. In other words, we examine the effectiveness of trading volume inflation in light of the conjectured tradeoff between short-term gains versus long-term losses due to harmed reputation. We obtain three interesting results. First, inflating trading volume by an exchange in a given currency pair in a given month increases the estimated real trading volume (i.e. trading volume controlling for fake trading estimate) in that pair over the course of next month – a one-standard-deviation increase in fake trading measures in a given month raises next month's real trading volume by 0.11-0.16 standard deviations. This suggests that trading volume inflation is beneficial to exchanges in the short run. Second, by augmenting the regressions of trading volume on lagged fake trading by lagged trading volume, we find that true trading volume is strongly positively associated with past volume and is negatively associated with past fake trading estimate. This result suggests that investors can partially see through exchanging washing trades. However, lagged trading volume is potentially endogenous. Thus, we instrument for lagged trading volume by its predicted value, which is strongly associated with lagged Bitcoin price. Our third finding is that when we use instrumented lagged trading volume the conclusion regarding investor sophistication, i.e. that investors can see through exchanges' trading volume manipulation, is reversed. The coefficient on lagged fake trading becomes insignificant, suggesting that investors do not distinguish between fake and real trading volume in the short run. This result is consistent with inability of market participants to decipher the quality of an experience good immediately upon purchasing it.

We then analyze longer-run effects of fake trading. In doing so, we recognize the possibility that the insignificant relation between past fake trading measures and current trading volume may be due to inability of our fake trading estimates to adequately measure current fake trading. Thus, we focus on non-volume-based outcomes, i.e. alternative measures of exchanges' operating success. The first such outcome variable is an exchange's web popularity, as measured by its Alexa rank. We find that the effect of fake trading on the web popularity of an exchange is positive and economically large at the medium-term horizon of three months: a one-standard-deviation increase in measures of fake trading leads to 1.5-1.9 standard-deviation increase in exchange's web popularity. Crucially, these effects are more than reversed at the longer-term horizon of twelve months: a one-standard-deviation increase in measures of fake trading is associated with 2.7-3.4 standard-deviation reduction in exchange's web popularity. These results are consistent with exchanges being able to mislead users in the short run but being punished in the long run.

Our second measure of operating success is an estimate of an exchange's revenues from commissions on the legitimate part of trading on its platform. To estimate the legitimate part of trading volume, we perform a counterfactual analysis of expected reduction in trading volume that would have occured had an exchange not inflated its volume. Our estimates suggest that on average 19% of reported volume is illegitimate, with wide variation across exchanges and currency pairs. Notably, these estimates are significantly lower than those in Cong et al. (2020), whose estimated mean proportion of fake trading volume is 70%.[8] This discrepancy in estimates may be due to at least two reasons. First, our estimates are likely downward-biased because they are relative in nature: We assume that that the benchmark exchange with the lowest fake trading estimate in a given currency pair does not inflate volume at all – an assumption that is likely optimistic. Second, our sample includes a wider range of currency pairs than previous studies, and we show that fake trading tends to be less prevalent in currency pairs involving alt coins, which constitute a majority of our sample but are not featured in Cong et al. (2020).

The results obtained using estimated-commission-revenue-based measure of operating success are largely consistent with the results for the web-popularity-based measure. The effect of fake trading on

[8] In addition, Aloosh and Li (2020) show that one third of trading volume in Bitcoin on Mt. Gox exchange was fake, and estimates in the Bitwise report (see Fusaro and Hougan (2019)) suggest that up to 95% of the overall Bitcoin trading volume is fake.

medium-term (three-months) estimated trading commission revenue is positive: a one-standard-deviation increase in measures of fake trading is associated with 0.4-0.5 standard deviation increase in estimated commission revenue. Longer-term (twelve-months) results indicate reversal: a one-standard-deviation increase in measures of fake trading leads to 0.2-0.3 standard deviation decrease in trading commission revenues. Overall, these results indicate that exchanges that fake trading volume tend to succeed initially in misleading investors but are usually punished eventually, consistent with the tradeoff involved in choosing an opportunistic strategy of volume inflation – between short-term (and medium-term) gains and long-term losses.

Our paper makes two main sets of contributions. Although we use a unique setting, our first set of contributions is quite general. First and foremost, we contribute to the literature on investment in reputational capital through product quality in the face of competition (e.g., Chamberlin (1933), Abbott (1955), Klein and Leffler (1981) and Shapiro (1983), Horner (2002), Dana and Fong (2011)). Although there is empirical literature that examines the relation between competition and product quality and generally finds that this relation is positive (e.g., Matsa (2011), Domberger and Sherr (1989), and Mazzeo (2003)), there is little empirical work that examines this relation in the context of Äúexperience goodsÄù, for which long-term reputation is key. One exception is Becker and Milbourn (2011), who focus on the credit rating industry and examine the effects of entry by Fitch on the quality of ratings produced by two incumbent rating agencies – Moody's and S&P.[9]

In addition to providing an out-of-sample corroboration of the result in Becker and Milbourn (2011) that increased competition leads to lower product quality, we contribute to the literature on product quality and reputation along the following dimensions. First, we examine the effects on product quality of both existing competitive landscape (static competition) and changes in it (dynamic competition). Second, the competitive landscape itself is more diverse in our setting, with multiple layers of competition – both in terms of currency-pair-level markets and in terms of geographical and operational aspects of competition at the exchange-level. Third, we examine the effectiveness of an opportunistic strategy of attempting to inflate the perception of product quality, i.e. we analyze whether fake trading has short-term benefits. Fourth, we analyze whether there is a tradeoff between short-term benefits and long-term reputational costs. Our findings that exchanges that inflate trading volume succeed in misleading investors in the short run but are punished in the long run are consistent with the model of Mailath and Samuelson (2001) in

[9]See also Mathis et al. (2009) for evidence of a negative relation between rating quality and expected costs of lost reputation.

which reputation is built gradually and dissipates gradually.

Second, our study contributes to the literature on competition and unethical behavior. Shleifer (2004) shows that unethical conduct is sometimes a consequence of product market competition. Consistent with Shleifer (2004), Luca and Zervas (2016) show that increased competition leads to more fake reviews of restaurants on Yelp platform. Bennett et al. (2013) demonstrate that competition decreases the quality of tests ("testing leniency") in the vehicle emissions testing market. Karpoff et al. (2008) study implications of announcements by regulators of discovery of financial misrepresentation ("cooking the books") and find that indirect penalties imposed by the market in the form of share price declines upon announcements dwarf the direct penalties imposed by the authorities. In our setting, characterized by generally lax regulation and enforcement, the extent of "cooking the order books" is never determined precisely and is gradually discovered by the market. Nevertheless, market participants are able to eventually see through much of the fake trading on exchanges and penalize them in the long run.

Our more specific contribution is to the nascent literature examining the behavior of crypto exchanges, which are one of the more important players in the crypto market (e.g., Amiram et al. (2020), Cong et al. (2020), and Griffin and Shams (2020)). We complement Cong et al. (2020), who perform the first large-scale analysis of fake trading on crypto exchanges, along the following dimensions. First, we corroborate their findings that a significant proportion of trading in four major crypto currencies against \$U.S. is likely fake. We show that these results continue to hold in a sample that includes almost the entire population of most important crypto currencies. Second, we analyze market-level, exchange-level, and currency pair-level characteristics that are associated with the extent of fake trading, with a focus on competition-related determinants of fake trading. Third, we examine implications of fake trading on future short-term and longer-term measures of exchanges' performance. Finally, in addition to using a measure of wash trading employed in existing studies, we introduce novel statistical measures of fake trading.

Our second and more specific contribution is to the nascent literature examining the behavior of crypto exchanges, which are one of the more important players in the crypto market (e.g., Amiram et al. (2020), Cong et al. (2020), and Griffin and Shams (2020)). We complement Cong et al. (2020), who perform the first large-scale analysis of fake trading on crypto exchanges, along the following dimensions. First, we corroborate their findings that a significant proportion of trading in four major crypto currencies against \$U.S. is likely fake. We show that these results continue to hold in a sample that includes almost the entire population of most important crypto currencies. Second, we analyze market-level, exchange-level, and currency pair-level characteristics that are associated with the extent of fake trading, with a focus on competition-related determinants of fake trading. Third, we examine implications of fake trading on future short-term and longer-term measures of exchanges' performance. Finally, in addition to using a measure of wash trading employed in existing studies, we introduce novel statistical measures of fake trading.

The remainder of the paper is organized as follows. The next section describes our data sources. In Section 3, we characterize main empirical regularities and the evolution of the crypto exchange industry over the last few years. In Section 4, we describe our measures of fake trading on crypto exchanges and perform several validation analyses. In Section 5, we analyze determinants of fake trading, focusing on factors related to static and dynamic competition among exchanges. Section 6 examines short-term and longer-term consequences of fake trading for exchanges' operating performance. Section 7 concludes.

2 Data

2.1 Main data source

Our main data source is Kaiko database (www.Kaiko.com), which provides details of every executed transaction on over 100 cryptographic exchanges, including currency pair, price, volume, and timestamp.[10] We focus on major exchanges, which we define as those that feature at least one currency pair in which the base currency is Bitcoin (BTC), Ether (ETH) or Tether (USDT). There are 41 of these, and they are responsible for 91% of the overall reported trading volume during our sample period.[11] The base currencies of these pairs have important functions in the crypto currency markets. Bitcoin (BTC) is the oldest crypto currency. Ether (ETH) is the base currency for the majority of tokens created in initial coin offerings (ICOs). Tether (USDT) is the main currency pegged to a fiat currency ($U.S.) and does not suffer from the high volatility of Bitcoin and most other crypto currencies.[12] The sample period begins in June 2013 and ends in September 2019. We validate our data by comparing it to data on www.CoinMarketCap.

[10]Kaiko collects, normalizes, and stores terabytes of historical crypto currency data from dozens of spot and derivatives exchanges. The procedure consists of retrieving the data either by using an exchange's API or by taking frequent snapshots of its web-based platform.

[11]We focus on this subset of exchanges and trading pairs in order to make the analysis operational. Even after imposing this filter, the total size of the data set is over 4.2 million files or 363 GB.

[12]To ensure that all relevant currency pairs are included in the analysis, we filter for inconsistencies in pair names. One type of inconsistency is pair ordering. For instance, in some exchanges the pair USDT-BTC is named as BTC-USDT. Another type of inconsistency is variations in the crypto currency ticker symbol. For example, Bitcoin Cash appears as BCH, BTCH or BTCASH on various exchanges.

com, which is the leading source of price and trading volume data on over 300 crypto exchanges, but unfortunately reports data at a low (daily) frequency. In contrast, Kaiko data is provided at a frequency of one second. As trades are not spread evenly over time, we aggregate all trades involving a given currency pair on a given exchange into ten-minute intervals.

2.2 Other data

We use numerous additional data sources that supplement Kaiko data. We obtain exchange characteristics, such as adoption of anti-money-laundering measures (AML), existence of know-your-customer requirement (KYC), and exchange location from www.CoinGecko.com and www.Cointelligence.com. Time series of web traffic data are obtained from www.Alexa.com. We collect time series data on technological development of exchanges, proxied by their code revisions on the most popular open-source platform, Github, from www.Github.com, and time-series evolution of exchanges' social media activity on two of the most popular social media platforms, Reddit and Twitter, from www.Reddit.com and www.Twitter.com, respectively. In separating crypto coins and tokens, we employ data from Lyandres et al. (2020), who collect token issuance data from 11 leading ICO aggregators and construct the most comprehensive data set of token issuance.[13]

3 The market

Panel A of Table 1 presents summary statistics of the cryptocurrency market over the duration of our sample (76 months).

[Insert Table 1 here]

The first row in Panel A of Table 1 presents aggregate crypto market capitalization, obtained from www.CoinMarketCap.com, which likely covers the whole population of crypto exchanges and currencies. The second row presents crypto market capitalization statistics aggregated within our Kaiko-based dataset. Our data covers a large fraction of the whole crypto market: During median month, 82% of the crypto market capitalization is included in our data, and this proportion is higher during the latter part of the sample,

[13]These aggregators are: www.Etherscan.io, www.CoinDesk.com, www.CoinGecko.com, www.CryptoCompare.com, www.ICObench.com, www.ICOdrops.com, www.ICOrating.com, www.ICOmarks.io, www.ICOdata.io, www.FoundICO.com, and www.Tokendata.io.

characterized by larger market size, as evident from the higher ratio of mean market capitalizations of currencies covered by the two data sets (87%).

Figure 1 presents the evolution of the crypto market capitalization over time.[14]

[Insert Figure 1 here]

The solid black line in Figure 1 depicts the market cap of Bitcoin, the dashed blue line refers to Ether, and the dotted red line represents combined market cap of all other crypto currencies ("alt coins"). The crypto currency market has grown tremendously between 2013 and 2019 – from 0.18 billion $US to over 800 billion $US at its peak, in January 2018, but since then saw a significant decline, which was mainly due to the decline in Bitcoin price and to "ICO winter", i.e. a sharp reduction in the volume of initial coin offerings since the middle of 2018.[15]

The mean number of distinct crypto currencies traded on exchanges in a given month, covered in our data, is 262, whereas at the end of our sample period, 967 currencies were traded on crypto exchanges. Over half of these are crypto tokens, which were issued via an ICO, whereas the rest are coins.[16] The mean number of new crypto currencies appearing on exchanges on a monthly basis is 33 and the mean number of currencies that exit the market in a month is 8; the medians are much lower, 2.5 and 0 respectively. The mean number of distinct crypto currency pairs traded on exchanges in a given month is 428, and at the end of our sample period, over 1,700 distinct pairs were trading on various exchanges; two thirds of these pairs involve crypto tokens. On average, 54 (10) currency pairs appear (disappear) in a given month. Figure 2 shows the evolution of entries of new currency pairs (solid blue line), exits of currency pairs (dashed red line) and net monthly change in the number of traded pairs across all exchanges (grey bars).[17]

[Insert Figure 2 here]

The mean reported aggregate monthly trading volume is $U.S. 38 billion or 6 million BTC. There are close to 100 million reported trades on average per month. Both the distribution of trading volume and

[14]For visualization purposes, in this and other figures, we omit data prior to 2016, when the rapid development of the crypto market started.

[15]In the year and a quarter since the end of our sample period, the crypto market capitalization rebounded and set a record of over $U.S trillion as of January 2021.

[16]While both tokens and coins are used to define a unit of value, there are important differences between them. First, tokens are built on top of existing blockchains, such as Ethereum, while coins, such as Bitcoin, Ether, or Litecoin, are unique digital currencies, which are based on their own, standalone blockchains. Second, and perhaps more importantly, coins are used mostly as a method of payment or a store of value, whereas tokens are also used to activate features on decentralized applications (DApps) they were designed for.

[17]In Figure 2, every appearance (disappearance) of a currency pair on (from) an exchange is counted, regardless of whether the currency pair is listed on other exchanges.

that of the number of trades are very skewed – the medians are one-to-two orders of magnitude lower than the means. Importantly, reported values of trading volume and the number of trades may be highly inflated due to abundance of fake (wash) trades on crypto exchanges, documented in Cong et al. (2020) — a finding that we corroborate and extend in this paper.

Panel A of Figure 3 presents the evolution over time of combined reported volume on all exchanges (in billion $U.S.), separated into the following subgroups of currency pairs: 1) ETH-BTC, 2) USDT-BTC, 3) BTC against other currencies, 4) ETH against other currencies, and 5) USTD against other currencies.

[Insert Figure 3 here]

The volume of trading in USDT saw rapid rise since its inception in late 2017, whereas the volume of trading in pairs involving ETH has been declining since its peak in the beginning of 2018. The peak of trading in BTC pairs preceded slightly the peak of the crypto market capitalization in January 2018. Panel B of Figure 3 presents the evolution of the number of trades (in million) across all exchanges for the same subgroups of currency pairs. The main patterns are similar to those evident in the evolution of trading volume. One interesting difference is that the reported number of trades in USDT pairs is low relative to reported volume, suggesting relatively large average trade size.

Crypto currencies are traded on multiple exchanges. The crypto exchange market (covered by Kaiko data and satisfying our selection criteria) grew from a single exchange – Bitfinex – in 2013 to 37 exchanges at the end of our sample period. The market is quite dynamic: on average, a new exchange is established every two months; exchanges do not seem to exit the market often – only four exchanges disappeared during our sample period, three of them in 2019.[18] The last three rows in Panel A of Table 1 describe the competitive landscape of the crypto exchange market, in particular the summary statistics of the market's Herfindahl indices, computed using three metrics: the number of currency pairs trading on an exchange in a given month, reported volume of trades on an exchange during a month, and reported number of trades on an exchange during a month. Figure 4 presents the evolution of the three HHI-based measures of crypto market concentration.

[Insert Figure 4 here]

[18]The frequency of exits may be understated, as Kaiko data focuses on a subset of most prominent exchanges, which have a lower likelihood of disappearance.

The market concentration has clearly decreased over time – from close to one throughout 2016 to 0.07–0.14 in late 2019, a decrease that coincides with the increase in the number of unique currency pairs — from 48 in early 2016 to over 1,700 at the end of our sample period.

Panel B of Table 1 presents statistics at the level of a crypto currency pair. Currency pairs tend to be traded on multiple exchanges. The mean number of exchanges on which a pair is listed is 2.46, and the median is 1.81; BTC-ETH pair was listed on 19 exchanges at its peak. The mean Herfindahl index of trading volume and the number of trades in a given currency pair across exchanges is around 0.25, suggesting substantial degree of competition among exchanges on average. The typical age of a crypto currency pair is 12 months on any exchange and 9 months on a given exchange; the BTC-LTC pair appears in every month throughout the sample. It takes a typical currency pair 6 months from initial listing on the first exchange to be listed on subsequent exchange(s).

Panel C of Table 1 reports summary statistics at the level of a crypto exchange. A typical exchange in our sample is 5 months old. The mean share of an exchange's reported volume (number of trades) out of the crypto market aggregate volume (number of trades) is 9.4%. The mean (median) number of currency pairs listed on an exchange in a given month is 78 (12).

The next few rows provide statistics of various characteristics of crypto exchanges that may be related to their optimal operating strategies. In 59% of exchange-months, an exchange has implemented anti-money-laundering measures (AML).[19] In 62% of exchange-months, an exchange has implemented know-your-customer (KYC) procedures.[20] 46% of exchanges are located in crypto-friendly coutries.[21] In 25% of exchange-months, there have been news on hack attacks, scams, theft, or poor review results associated with the exchange (Bad News).[22] In 12% of exchange-months, an exchange operates multiple platforms

[19]AML indicator equals one if the exchange has implemented an AML policy and provides detailed information about conformity with accepted international AML procedures. We use trust score from www.CoinGecko.com for all exchange characteristics discussed in this paragraph.

[20]KYC indicator equals one if there is evidence that the exchange provides clear guidelines as to the documents required for verification of the sources of clients' funds.

[21]The list of crypto friendly countries includes Singapore, Russia, Estonia, Malta, Luxembourg and Switzerland. A country is considered crypto-friendly if there is evidence that established exchanges and crypto-businesses move to it due to friendly regulatory environment, e.g., wide acceptance of crypto currencies, positive regulatory developments, and/or existing guidelines and easy registration with local financial authorities.

[22]News about hacks or crypto exchange misuse of funds are often reported in popular outlets, e.g., https://www.theguardian.com/technology/2019/jul/12/tokyo-cryptocurrency-exchange-hack-bitpoint-bitcoin. According to www.CoinTelegraph.com, the total amount of funds stolen from crypto exchanges as of November 2020 is $U.S. 2.8 billion.

(Multiplatform).[23]

The next four rows in Panel C describe measures of exchanges' popularity, their social media presence, technological advancement, and transparency. Alexa captures the mean popularity rank of an exchange's website (out of all contemporaneous sites on the world wide web), with the rank of one corresponding to the most popular website, available from www.Alexa.com.[24] The most popular exchange is Binance, whose website is ranked 230 on average. The mean (median) rank or a crypto exchange in our sample is 41,750 (9,470). For comparison, the average ranks of some of the main traditional exchanges — NASDAQ, NYSE, and Euronext are approximately 3,800, 27,000, and 62,000, respectively, as of November 2020. Reddit and Twitter measure social media activity related to the exchange – discussions on Reddit and posts on Twitter under firm's handle. The mean monthly numbers of Reddit posts and Twitter tweets by an exchange are 38 and 178, respectively, whereas median exchange is not active on social networks. Github is the number of commits (code reviews) of an exchange's platform, proxying for its technological advancement.[25] A median exchange sees 13 revisions to its code in a month, whereas the mean monthly number of code revisions is over 100. Since providing an open-source code is optional, the existence of code revision on an open-source platform is an indication of transparency of an exchange. The last few rows in Panel C present the distribution of exchange locations, aggregated into geographical regions.[26] Almost 40% of exchange-month observations are in Asia (including China) and only 20% of exchanges are in Western Europe or North America.

Panel D of Table 1 presents summary statistics at the level of a currency pair traded on a given exchange. The mean reported volume of trading of a currency pair on a single crypto exchange is $U.S. 46 million per month, whereas the median is lower than $U.S. one million. This skewness is driven by several abnormal volume numbers. For example, the highest reported monthly volume was $U.S. 49 billion in March 2019 for Peercoin-Bitcoin pair (PPC-BTC) at the BX.in.th exchange – once Thailand's largest crypto exchange, which by now has ceased operations. Similarly, the mean number of trades in a currency

[23]For example, Binance has a regular, centralized, platform, but also a decentralized platform, since April 2019 (see https://www.coindesk.com/binance-launches-decentralized-exchange-ahead-of-schedule). One reason exchanges provide users access to alternative platforms is an attempt to guard their market share in wake of the rise of decentralized exchanges (DEX), which reduce risks to users of misuse of funds and eliminate the need in a market maker.

[24]Alexa rank is computed using a methodology proprietary to www.Amazon.com that combines a site's estimated traffic and visitor engagement over the past three months.

[25]We only count commits in the main exchange repository when they are available.

[26]In many cases the precise location of an exchange is unclear and a region-level measure (as opposed to a country-level one) reduces the measurement error. For instance, Okex website states that the exchange is based in Belize, but its main operations are located in Hong Kong and the headquarters are in Malta. In cases involving discrepancies, we base the assignment of an exchange to a geographical region based on the location of its main operations.

pair on a given exchange is over 120,000, whereas the median number is an order of magnitude lower. This difference is also driven by an unusual number of trades in some pairs on particular exchanges. For example, there were over 25 million trades in Aeternity-Bitcoin pair (AE-BTC) at the ZB exchange in March 2019. As these examples suggest, there is a strong possibility that some of these outlier values of trading volume and the number of trades are driven by fake (wash) trading – a hypothesis we examine in detail below. Trading on some exchanges is heavily concentrated in a given currency pair – the highest Herfindahl index of trading volume of currency pairs on a given exchange is one, while some exchanges are well diversified – the lowest volume-based HHI is 0.02. The mean and median volume-based and number-of-trades-based Herfindahl indices on a given exchange equal 0.5, suggesting significant reliance of exchanges on the most popular pairs traded on them.

4 Measuring fake trading

As the focus of this paper is identifying determinants and examining consequences of fake (wash) trading on crypto exchanges, we begin by providing several examples that illustrate time series patterns of trading activity that may be suggestive of wash trading. We then proceed to developing quantitative measures of fake trading that we use in the empirical analysis.

4.1 Illustrations of (potential) fake trading

Figure 5 presents several examples of trading activity that looks unusual and may be indicative of fake trading.

[Insert Figure 5 here]

Panel A depicts the evolution of trading volume (in the upper two figures) and the number of trades (in the lower two figures) in the OmiseGo-Bitcoin pair (OMG-BTC) during 4,464 ten-minute intervals in January, 2019 on two exchanges – Okex and Binance. The upper two figures suggest that while trading volume on Binance follows an arguably random pattern, there seem to be several structural breaks in trading volume on Okex. Similar structural breaks are evident in the number of trades on Okex, whereas they do not seem to be present on Binance.

Panel B presents trading volume and number of trades in one of the most important crypto currency pairs — ETH-BTC — on two exchanges: ZB (on the left) and Binance (on the right) in April, 2019. While the plots of both trading volume and the number of trades on Binance do not reveal unusual patterns, there are numerous spikes of trading activity on ZB, which are often two orders of magnitude larger than typical trading volume, which occur with a roughly constant frequency.

Panel C depicts trading volume and number of trades in Time New Bank-Bitcoin pair (TNB-BTC) pair on Huobi and Binance. Once again, the trading volume on Binance does not suggest specific patterns. In contrast, there are several peculiarities in the graphs of trading volume and the number of trades on Huobi. First, both trading volume and the number of trades is constant for prolonged periods of time. Second, there are clear structural breaks in the number of trades plot. Finally, the average trade size on Huobi is about three orders of magnitudes smaller than that on Binance.

While these are just suggestive examples and are very far from serving as a smoking gun in uncovering wash trading, they guide us in developing quantitative measures of potentially fake trading. In what follows, we attempt to design measures that would uncover irregularities/abnormalities similar to those in Figure 5 and others, which are less visible to a naked eye, and may be suggestive of fake trading.

4.2 Measures of fake trading

Measuring fake trading is difficult and no direct measures exist. According to Gerald Chee, head of research at `www.CoinMarketCap.com`, "there is no way to tell if an exchange is inflating volume or not by merely looking at the volume they report. The only way to detect 'wash trades' would require access to 'account-ID' data... only exchanges have access to these [data]". In the absence of direct measures of fake trading, we use indirect, statistical measures developed both in the prior literature and here, which we describe in this section.

4.2.1 Benford's law

The first measure of irregular trading patterns that we use applies the Benford's Law (Benford (1938)), which is based on the likelihood of occurrences of first digits of a series. In many naturally observed series, which are not constrained to a certain range, 1 appears as the leading digit 30.1% of cases. As the value of the first digit increases, its frequency decreases: 2 (3, 4, 5, 6, 7, 8 and 9) is the leading digit in 17.6% (12.5%, 9.7%, 7.9%, 6.7%, 5.8%, 5.1% and 4.6%) of the time. Deviations from Benford's Law may indicate abnormalities in data series.

Panel A of Figure 6 shows examples of observed series of first digits of trading volume during ten-minute intervals over the course of one month (March 2019) in the ETH-BTC pair on four crypto exchanges – Binance, ZB, Okex, and Bibox.

[Insert Figure 6 here]

In each of the four figures, the solid line depicts the actual distribution of leading digits of the number of trades, whereas the dashed line depicts the theoretical, Benford-Law-based distribution. The observed series for Binance seems to roughly approximate the theoretical Benford's Law. However, the other plots indicate frequencies of leading digits that deviate substantially from Benford's Law. For instance, the plot for ZB shows that the frequencies of digits 7-9 are over twice as large as those predicted by Benford's Law; the same is true for digit 5 on Okex and digits 2 and 3 on Bibox. The deviations from Bedford's Law are even more striking in Panel B of Figure 6, which presents examples of frequencies of first digits of the number of trades series.

To quantify the deviations from Benford's Law, we calculate the mean absolute deviation (MAD) between the theoretical Benford's Law-based fractions of leading digits and the proportions in the observed series. MAD is calculated as the equally-weighed average of the absolute distances between the observed value and the Benford's-Law-based value across the nine digits (e.g., Drake and Nigrini (2000)).[27] Summary statistics of MAD, reported in the first two rows of Panel A of Table 2 indicate that there is wide variation in MAD for both trading volume and the number of trades.

[Insert Table 2 here]

4.2.2 Distance from log-normal distribution

Empirically, in the absence of evident manipulation, trading volume tends to follow log-normal distribution (e.g., Richardson et al. (1986) and Ajinkya and Jain (1989)). Sizable deviations from log-normality may be indicative of irregularities in/manipulation of trading. A convenient measure that quantifies the distance between a sample's empirical cumulative distribution function (c.d.f.) and the c.d.f. of a reference distribution is Kolmogorov-Smirnov (KS) statistic. KS distance is given by the supremum of the distance between the c.d.f. of the theoretical distribution and that of the observed one over all realizations of the

[27] Our results are robust to employing another measure of deviations from Benford's Law — the standard absolute deviation (SAD) — which equals the standard deviation of the absolute differences between the theoretical, Benford-Law-implied, and observed series.

variable (e.g., Conover (1971)). Thus, we compute the KS distance between the c.d.f. of the empirical distribution of the natural logarithm of reported trading volume and the c.d.f. of normal distribution with the same mean and variance. In particular, for each exchange-month and for every currency pair, we first compute the mean and standard deviation of log trading volume over ten-minute intervals. We then compute the maximum distance between the c.d.f. of the resulting empirical distribution and that of normal distribution with the same mean and variance. We perform this exercise for every exchange-month-currency pair and repeat the procedure while substituting trading volume by the number of trades.

Figure 7 illustrates the c.d.f. of the natural logarithm of trading volume (solid line) on four exchanges – Binance, Okex, ZB, and Bibox and the c.d.f. of normal distribution with the same mean and variance (dotted line) – for ETH-BTC pair in March, 2019 – for trading volume in Panel A and for the number of trades in Panel B.

[Insert Figure 7 here]

The c.d.f. of the volume of the log of trading on Binance and Okex are the closest to normal: the Kolmogorov-Smirnov distances are 0.14 and 0.18, respectively. On the other end of the spectrum is ZB, with KS statistic of 0.73, whereas Bibox is in between (KS statistic of 0.30). Similar results are obtained for number-of-trades-based KS distances: The lowest are those for Binance and Okex (0.12 and 0.19, respectively), whereas the highest is for ZB (0.54). Summary statistics of KS distance, found in the third and fourth rows of Table 2, indicate a roughly symmetrical distribution centered around 0.3.

4.2.3 E-divisive with medians

Another method that we use for flagging unusual patterns in trading is rooted in the computer science literature that focuses on detecting structural breaks, i.e. mean shifts or trend shifts in a series. One particular algorithm that is suitable for detecting structural breaks in trading volume and the number of trades is called EDM (E-Divisive with medians (e.g., James et al. (2016)). The algorithm attempts to determine, through comparisons of various permutations of the data, whether a new chunk of time series data is considerably different from the previous chunk.[28]

One important reason for the suitability of EDM in our setting is that it is robust to the presence of short-lived abnormalities in the series. This is particularly relevant because trading volume data tend

[28]EDM is frequently used for detecting breakouts in data transmission in e.g., mobile internet applications.

to have occurrences of peaks, which, despite often being a result of a natural data generating process, may appear as anomalies when data series is aggregated into short intervals. In addition, EDM is non-parametric. This is important because trading volume that is partially a result of wash trading does not usually conform to a particular distribution. The trading volume and number of trades series often contain more than one structural break. Our EDM-based measure of fake trading is the number of breaks within a monthly series.

Figure 8 presents two examples of the performance of EDM algorithm in detecting structural breaks in the data.

[Insert Figure 8 here]

The first example, presented in Panel A, is the number of trades within ten-minute intervals in the OMG-BTC pair on Okex exchange in January 2019. The series is characterised by a step-like shape. The EDM algorithm is able to correctly detect four structural breaks, highlighted by dashed vertical lines (where the timing of the breaks is identified by the algorithm). Importantly, due to its reliance on comparisons of medians (and not means), the algorithm also correctly ignores a temporary spike in trading around bin 3,800. The second, example, depicted in Panel B, is the number of trades in the ETH-BTC pair on Binance exchange in May 2019. The EDM algorithm does not identify any breakouts in this case, consistent with no visible structural breaks in the trading volume plot. We used numerous examples such as the ones in Figure 8 to calibrate the parameters of the algorithm, such as the minimum size of buckets, and penalization parameters.[29]

4.3 Aggregating measures of fake trading

None of the three measures of fake trading described above are measuring without errors the underlying construct of fake trading. Each is constructed to identify particular deviations from normal trading, frequently observed in the data. Thus, in order to construct a unified measure of fake trading, we extract principal components of the measures described above at the exchange-month-currency pair level.[30] To

[29]The minimum size of buckets is the lowest number of observations between structural breaks, where each observation is the trading volume or the number of trades within a ten-minute interval. We set the minimum bucket size at 6. We follow James et al. (2016) in choosing parameters of polynomial penalization, which determine the sensitivity of the model. In particular, degree of penalization is 1 and β equals 0.008. Our results are robust to various modifications of parameter values. For more details about the EDM algorithm see `https://github.com/twitter/BreakoutDetection`.

[30]Most of the empirical results presented below are robust to using individual measures of fake trading, based on Benford's Law, Kolmogorov-Smirnov distance from log-normal distribution, and the number of structural breaks identified by the EDM algorithm.

prevent dominance of particular measures due to their high variances, we scale all measures of fake trading to have zero mean and unit standard-deviation.

Figure 9 presents fractions of variance explained by each principal component – for principal components derived from the trading-volume-based measures in Panel A, for those derived from the number-of-trades-based measures in Panel B, and for those derived from both types of measures in Panel C.

[Insert Figure 9 here]

Since the first principal components explain significant fractions of variation in fake trading measures — 52-63%, we concentrate on the first principal components as aggregated measures of fake trading in the empirical analysis.

Figure 10 presents biplots that show the orthogonalization of the measures of fake trading and their relation with the first two principal components (for trading-volume-based, number-of-trades-based and both volume-based and number-of-trades-based measures in Panels A, B, and C, respectively). The length of each vector (squared cosine) represents the respective variable in the first two principal components. The horizontal and vertical projections of a vector represent the variable in the first and second principal component respectively.

[Insert Figure 10 here]

As evident from the three panels, all measures of fake trading are positively correlated with the three versions of the first principal component, which simplifies the interpretation of the signs of the relations between the principal-component-based fake trading measures and the variables of interest in the analysis below. The correlations between individual measures of fake trading and their first principal components range between 0.6 and 0.77.

Figure 11 (A) presents mean principal-component-based measures of fake trading over time. In particular, every quarter, we compute equally-weighed average over all exchange-currency pair-months of the three versions of the first principal component of fake trading. The horizontal lines centered around point estimates depict confidence intervals of these estimates.

[Insert Figure 11 (A) here]

Mean measures of fake trading are generally increasing since 2017. Temporal increase in fake trading – an interesting finding in itw own right – is potentially consistent with the entry of less reputable exchanges, with fake trading being more prevalent in newer, less liquid currency pairs, and with the temporal increase in the intensity of competition among exchanges. We examine these conjectures below.

Figure 11 (B) depicts mean measures of fake trading over time, separated into five groups of currency-pair types: 1) ETH-BTC, 2) USDT-BTC, 3) BTC against other currencies, 4) ETH against other currencies, and 5) USTD against other currencies.

[Insert Figure 11 (B) here]

The measure depicted in this Figure is the first principal component of both volume-based and number-of-trades-based measures.[31] The lowest level of fake trading is observed in pairs involving BTC and ETH on one side and alt coins on the other side, perhaps because these pairs tend to be rather illiquid on average and detecting fake trading in such pairs is the easiest. Most fake trading occurs in the USDT-BTC pair, followed by the ETH-BTC pair, which are, as of late, the pairs with the highest volume of trading.

4.4 Validation of fake trading measures

As our measures of fake trading are purely statistical and are partially based on machine learning, it is important to examine their validity, i.e. to verify that they are correlated with exchange and/or currency pair characteristics that are likely to be associated with quality. We present four types of suggestive evidence of the validity of our fake trading measures. First, we show that our estimates of fake trading are lower for most established and popular exchanges than for newer and less popular ones. Second, we demonstrate that exchanges that are subject to stricter regulation/compliance tend to have lower estimates of fake trading. Third, we show that our measures are correlated with more intuitive but less precise measures of fake trading, which are based on correlations of trading volume and number of trades on a given exchange with aggregate contemporaneous trading volume and number of trades. Fourth, we present evidence based on a quasi-natural experiment – the ban by China of some exchanges that used to be based there – that shows that our measures of fake trading change in the expected direction following this policy shock.

4.4.1 Fake trading estimates by exchange

Figure 11 (C) presents exchange-level averages of our three fake trading measures.

[31] The results are similar for volume-based principal component measure and for number-of-trades-based one.

[Insert Figure11 (C) here]

Exchanges with the highest levels of web popularity, among those covered in our data – Binance, Coinbase, and Upbit, as well as the oldest exchanges – Kraken, Bitfinex, and CEX.IO, all have below-sample-mean measures of fake trading. In addition, there exchanges in our sample that are regulated by the New York State Department of Financial Services – Coinbase, Bitflyer, and Gemini – have some of the lowest fake trading measures. On the other hand, some of the exchanges with the highest estimated fake trading measures – CoinEx, Bibox, Okex, and Huobi – have been subject to reporting that may suggest subpar governance.[32]

4.4.2 Fake trading estimates and regulation/compliance/transparency

In Table 3 we compare our measures of fake trading between groups of firms with relatively high and low levels of self-imposed and external regulation and compliance, as well as high and low levels of transparency. For each of the three principal-components-based measures of fake trading, we report mean measure for groups of exchanges with high/low values of characteristics potentially associated with regulation/compliance and the difference between the two subsets. Most of the variables correlated with regulation/compliance are discrete. For continuous variables, groups are formed using the median of the variable values.

[Insert Table 3 here]

Exchanges with anti-money-laundering (AML) provisions in place have lower mean fake trading measures than those without such provisions, the difference in means being highly significant for all three principal-component-based fake trading measures. Similarly, exchanges with know-your-customer (KYC) requirements have significantly lower fake trading measures than exchanges without such requirements. Exchanges located in countries with crypto-friendly regulations and exchanges subject to negative news coverage are characterized by significantly higher fake trading measures than exchanges located in countries with stricter regulation and those without bad publicity. Exchanges operating both centralized and

[32]See: https://www.trustpilot.com/review/coinex.com for examples of user complaints regarding CoinEx. See complaints about possible manipulation of Bibox and Okex trading volumes at https://medium.com/@pandahanda44/bibox-artificially-inflated-volume-in-pictures-ebf6cbe1ed2a and https://medium.com/@sylvainartplayribes/chasing-fake-volume-a-crypto-plague-ea1a3c1e0b5e, respectively. See complaints about funds disappearance on Huobi: https://medium.com/@revblc/hacked-at-huobi-or-an-inside-job-you-decide-7979553cae9f.

decentralized platforms tend to have higher fake trading estimates (on their centralized platform). A possible reason is that washing trades on a decentralized platform would be easily detectable and, thus, if an exchange is interested in inflating volume, it would likely do so on its centralized platform. Finally, exchanges that are less popular (based on their Alexa rank), those that have relatively low social media presence (on Twitter and Reddit), and exchanges that are more technologically opaque (i.e. those without Github commits) tend to have higher fake trading estimates.

4.4.3 Alternative fake trading measures

Table 4 examines the relation between our principal-components-based measures of fake trading with alternative measures, which are based on correlations between either the number of trades or trading volume in a given currency pair within a ten-minute interval on a given exchange with the same variable aggregated for that currency pair on all exchanges. The idea behind the correlation-based measures is simple. As shown in Figure 5, suspicious trading activity can take many shapes. Assume that on each exchange, trading consists of two components. The first is the legitimate one, driven by information/news, that is likely positively correlated with contemporaneous legitimate trading in the same currency pair on other exchanges. The second is fake trading, which, given its various shapes and frequencies, is generally uncorrelated across exchanges. Intuitively, trading volume and the number of trades within a given time interval on an exchange with a less significant fake trading component is likely to have a higher correlation with the aggregate contemporaneous trading volume or the number of trades than this correlation of an exchange with a larger fake trading component.[33]

To estimate the relation between our measures of fake trading on one hand and correlations between trading volume or number of trades on an exchange and aggregate quantities on the other hand, we follow the following steps. First, each month for each exchange and currency pair that is traded on at least five exchanges, we compute the correlation between the volume of trading on that exchange during ten-minute intervals and aggregate volume of trading on all exchanges during the same intervals. We repeat this procedure for the number of trades. Second, we regress the estimated monthly correlation for a given currency pair on a given exchange on the contemporaneous estimate of one of the three principal-component measures. Given that the correlation between trading volume on an exchange and aggregate volume is likely related to the popularity of the exchange — volume of trading on more popular exchanges

[33]Simulations in which we estimate these correlations while varying the proportion of fake trading on various exchanges are consistent with the intuitive argument.

is likely to exhibit a more positive correlation with the aggregate volume — in some specifications we control for the exchange's web popularity and for its social media activity.

[Insert Table 4 here]

The upper part of the first column in Table 4 shows that the correlation of trading volume on a given exchange with aggregate trading volume is significantly negatively related to our volume-based principal component measure of fake trading. The second column demonstrates that the negative relation remains significant, albeit weaker, when the exchange's web popularity and social media presence are controlled for. The third and forth column show similar results when the correlation of trading volume is replaced by the correlation between the number of trades on an exchange and contemporaneous aggregate number of trades. Similar results are obtained when instead of volume-based principal component, we use number-of-trades-based principal component (in the middle part of the table) and the principal component based on both trading volume and the number of trades (in the bottom part of the table).

In the last two columns, we perform a counterfactual analysis to ensure that the negative coefficients on principal-components-based measures of fake trading that we report in columns 1-4 are not mechanical. We use the correlation between the average exchange rate of a currency pair across all transactions on a given exchange within ten-minute intervals and the average exchange rate of that currency pair across all transactions on all exchanges during the same intervals. In other words, we replace the volume-based and number-of-trades-based correlations by the price-based correlation. The idea is that wash trades are typically not performed to systematically influence prices but rather to inflate trading volumes. Thus, we expect no relation between price-based correlation and principal-components-based measures of fake trading. The results are consistent with this conjecture: This relation is small and statistically insignificant in all six specifications in columns 5-6 of Table 4. Overall, the results in Table 4 suggest that our statistical measures of fake trading are correlated with alternative, more intuitive, correlation-based measures, providing further evidence of the validity of our measures. Importantly, our results are robust to including these correlation-based measures into principal-components-based aggregates.

4.4.4 Fake trading estimates and the Chinese ban

In Table 5, we analyze the effects of a ban on crypto exchanges imposed by regulators in China in 2017 on fake trading estimates. In early 2017, leading Chinese exchanges at the time — Huobi, BTCC, and OkCoin — were responsible for over 90% of total worldwide volume of Bitcoin trading. The regulatory action in September 2017 consisted of banning initial coin offerings, restricting any activity that involves crypto tokens, shutdown of exchanges' websites, and canceling exchange executives' business licences.[34] As a result, several Chinese exchanges — BigONE, EXX, CoinEx, Gatecoin, Huobi, OkCoin and OkEx — moved operations outside of Mainland China to Hong Kong, Singapore, and countries with lax regulation, such as Malta.

[Insert Table 5 here]

We use the Chinese ban as an exogenous shock to the propensity of inflating volume for (initially) Chinese exchanges. Relative to pre-ban China, Hong Kong's and Singapore's regulatory environments are strict, making wash trading more difficult. Thus, we expect a reduction in the extent of fake trading in exchanges that moved from China to other Asian locations as a result of the ban. We use this event to further validate our measures of fake trading by examining whether there are significant changes in estimated fake trading on these exchanges following the ban.

To examine the effects of the ban on fake trading estimates, we construct two samples. In the first one, the treated group includes seven aforementioned Chinese exchanges that relocated outside of China as a result of the ban. The control sample includes exchanges that were operating in Hong Kong, Singapore, Korea and Japan at the time of the ban – Binance, Bitfinex, ZB, Bit-Z and Ethfinex, Kucoin, BitForex, UPbit, BitBank, Quoine and Zaif. In the second sample, we restrict the set of treated exchanges to those that moved from mainland China to Hong Kong – OkCoin, BigONE, CoinEX, EXX, and Gatecoin. The control sample includes five exchanges that were operating out of Hong Kong at the time of the ban and remained there following the ban – Binance, Bitfinex, ZB, Bit-Z and Ethfinex.[35]

Our sample period for this analysis includes the first three quarters of 2017 (the "pre-ban period") and the first two quarters of 2018 (the "post-ban period"). We construct two indicator variables: the first takes the value of one for the treated exchanges ("Treated"); and the second indicator equals one for the post-ban period ("Post-ban"). We then estimate regressions of our three principal-components-based measures of fake trading at the exchange-month-currency pair level on the treated indicator, the post-ban indicator, and the interaction between the two, while controlling for all exchange-level and currency-pair-level characteristics that we describe in detail below, in Table 6.

[34] See https://www.coindesk.com/chinas-ico-ban-a-full-translation-of-regulator-remarks for details of the regulation.

[35] We rely on www.CoinGecko.com and www.Cointelligence.com for obtaining historical locations of each exchange in our dataset.

The coefficient on Treated is positive and highly economically and statistically significant in all regressions – the level of fake trading on Chinese exchanges prior to the the ban was 0.8-2 standard deviations higher than on exchanges located in Pacific Asia/Hong Kong. (In all regressions here and below, all explanatory variables are normalized to have zero mean and unit standard deviation, to ease the interpretation of the coefficients' economic significance.) The negative coefficients on Post-ban are also large and significant, indicating that all exchanges reduced their fake trading in the first two quarters of 2018 relative to the first three quarters of 2017. Most importantly, the coefficients on the Treated × Post-ban interaction are negative and highly statistically significant and economically large — Chinese exchanges that moved out of China following the ban reduced the extent of their fake trading by 0.7-1.3 standard deviations compared to their local peers.

5 Determinants of fake trading

5.1 Static competition among exchanges and fake trading

In Table 6, we examine associations between exchange and currency pair characteristics on one hand and our measures of fake trading on the other hand.

[Insert Table 6 here]

In odd columns, we include both exchange-level and currency-pair level characteristics, whereas in even columns, we focus on currency-pair-level ones, while including exchange fixed effects. In all regressions, we control for the base pair fixed effects (i.e. whether the base pair is BTC, ETH, or USDT) and for time (year-quarter) fixed effects.

Exchanges on which more currency pairs are traded tend to have higher measures of fake trading: A one-standard-deviation increase in log of the number of trading pairs is associated with 0.04-0.16 standard-deviation increase in fake trading. This is consistent with the results in Dana and Fong (2011), who show that maintaining reputation is easier in the presence of multimarket competition. Larger exchanges (measured by the overall share of an exchange's reported volume out of total aggregate volume) tend to have lower estimates of fake trading. Interestingly, this negative relation is observed despite the positive mechanical link between wash trading and volume, since the volume of wash trading is part of the overall reported volume. Exchange age, on the other hand, is not an important determinant of the extent of fake trading. While it is significantly negatively related to measures of fake trading in two out of three specifications, the economic magnitude of this relation is quite small – a one-standard-deviation of log(age) is associated with 0-0.04 standard-deviation reduction in the extent of fake trading.

Using exchanges located in North America as a benchmark, our results indicate that exchanges in Asia (and especially in China) and in European jurisdictions with lax regulatory oversight ("Europe: Islands"), as well as those in Central and South America and Africa tend to be characterized by high levels of fake trading. The results for exchanges located in Western and Eastern Europe are less conclusive: average fake trading measures there tend to be roughly on par with those in North America.

Exchanges do not fake trading equally in all trading pairs. There is a positive relation between the number of exchanges on which a pair is listed and the average level of fake trading in the pair. This is consistent with concerns about future reputation damages being lower in highly competitive markets. In line with this interpretation, concentration of trading in a given pair across exchanges is negatively related to the degree of fake trading – lowering the Herfindahl index based on reported trading volume across exchanges by one standard deviation is associated with 0.12-0.22 standard-deviation increase in our fake trading estimates.[36] Exchanges do not significantly change the extent of volume inflation as currency pairs mature, as evident from the largely insignificant relation between the age of a pair and fake trading measures.

There is significantly less fake trading in currency pairs involving tokens issued in an ICO. A possible reason is the difference between exchanges' incentive structures for tokens and for coins. The bulk of compensation of an exchange from listing a token comes at the listing stage, in the form of fixed listing fee. On the contrary, most of an exchange's profits from listing pairs involving coins comes from trading commissions, which raises incentives to signal quality in coin-based pairs by inflating reported trading volume and the number of trades.

5.2 Dynamic competition among exchanges and fake trading

The results in the previous subsection suggest that static measures of competition, such as size, diversification of an exchange across currency pairs, and concentration of trading in a given currency pair across exchanges are related to measures of fake trading. We now examine how dynamic aspects of competition among exchanges influence the extent of volume inflation. In Table 7 we examine the effects of

[36] Similar results are obtained when volume-based Herfindahl index is replaced by number-of-trades-based Herfindahl index.

entry and exit of rival exchanges on fake trading measures on a focal exchange, while controlling for all exchange-level and currency-pair-level characteristics highlighted in Table 6, as well as base pair and time fixed effects.

[Insert Table 7 here]

Entry of an exchange into a given currency pair is a situation in which an exchange that has not previously listed the currency pair reports trading in it in a given month. Exit of an exchange is a situation in which the exchange that has listed the currency pair last month does not list it in a given month. We define three types of competitors. "General" competitors are a set of all exchanges in our dataset. "Geographical" competitors are a subset of exchanges that operate in the same geographical region as the focal exchange, where regions are defined as in Panel C of Table 1. Geographical dimension of competition is important in the crypto market. Shams (2020) shows that the clientele of investors on crypto exchanges is strongly related to their location. "Operational" competitor is one exchange belonging to the set of general competitors that has the largest overlap of pairs listed with the focal exchange.[37] Operating competitors' entry/exit may have especially pronounced effects on the focal exchanges.

We include two measures of static competition, captured by indicator variables – "Moderate competition", which equals one for a currency pair that is listed on at least two and at most seven exchanges, and "High competition", which equals one for currency pairs listed on at least eight exchanges. Consistent with the results of the effects of static competition on fake trading in Table 6, all three fake trading measures are monotonically increasing in the extent of competition. The coefficients on the moderate competition dummy are positive and significant and the coefficients on the high competition indicator are also highly significant and are larger in magnitude: Currency pairs that are traded on at least eight exchanges have 0.4-0.6 standard-deviation higher fake trading measures on average than currency pairs traded on a single exchange.

An (additional) exchange listing a currency pair ("competitor entry") is associated with 0.14-0.22 standard-deviation increase in measures of fake trading in that pair on the focal exchange. This increase tends to be somewhat larger if the rival exchange is operating in the same geographical region or if the rival exchange has a large overlap with the focal exchange in the set of currency pairs listed: in these

[37] The overlap is computed as the ratio of the number of pairs listed on both exchanges to the number of distinct pairs listed on at least one of the two exchanges. Another possible way to define operational competitors is by performing textual analysis of crypto news, as in Schwenkler and Zheng (2019) and Schwenkler and Zheng (2020)).

cases the increase in fake trading measures is 0.17-0.27 standard deviations and 0.20-0.29 standard deviations respectively. When a currency pair listed on an exchange is being delisted on another exchange ("competitor exit"), there is a reduction in fake trading estimates on the focal exchange. This result holds for the full set of competitors; however when subsets of competitors are restricted based on geography or operational overlap, the relation between a competing exchange's exit and our fake trading measures is insignificant. In addition, the effects of competitor entry and exit tend to be more pronounced when the extent of competition is high (i.e. for currency pairs that are listed on many exchanges). The relation between the interaction of competitor entry and high competition on one hand and fake trading on the other hand is positive and significant in six specifications out of nine. The relation between the interaction of competitor exit and high competition on one hand and fake trading on the other hand is significantly negative in five specifications.

6 Consequences of fake trading

Results in the previous section suggest that fake trading may be partially driven by the competitive pressure leading exchanges to inflate trading volume. In this section we examine short-term and long-term effects of fake trading.

6.1 Effects of fake trading on trading volume

We begin by asking a question: Is fake trading effective? In other words, while trading volume inflation mechanically increases overall contemporaneous volume, an interesting question is: Does volume inflation on an exchange in a given period have implications for the (legitimate) trading activity on that exchange in the future. To answer this question, we estimate regressions at the exchange-currency pair-month level of trading volume on lagged measures of fake trading, while controlling for current fake trading and for exchange and currency pair characteristics, as in Tables 6 and 7.

[Insert Table 8 here]

Trading volume is mechanically positively related to contemporaneous volume-based measure of fake trading, as evident from the first column of Table 8. More interestingly, lagged measure of fake trading is positively related to current volume: A one-standard-deviation increase in last month's fake trading measure is associated with 0.11 standard-deviation increase in current volume. Similar results are obtained in columns 5 and 9, in which the volume-based fake trading measure is replaced by the number-of-trades-based measure and a measure based on both trading volume and the number of trades, respectively.

The likely mechanism behind the effect of lagged fake trading measures on contemporaneous trading volume is through inflated lagged volume that wash trading causes. Thus, in columns 2, 6, and 10, we augment the regressions by lagged trading volume. Lagged volume has a profound effect on current volume: A one-standard-deviation increase in last month's volume is associated with a 0.9 standard-deviation increase in current volume. This is consistent with strong autocorrelation in trading volume at the exchange-currency pair level that is present in the data. Importantly, after augmenting the regression by lagged trading volume, the coefficients on lagged fake trading measures flip sign and become negative. Conditional on overall (legitimate and fake) lagged trading volume, a one-standard-deviation increase in fake trading is associated with 0.05-0.08 standard-deviation decline in current trading volume. The interpretation of this result is that market participants can, to a certain degree, distinguish between legitimate and fake trading volumes. However, they do not fully internalize the extent of fake trading, as inflating trading volume still has some positive impact on future volume, as evident from columns 1, 5, and 9.

Importantly, the estimates discussed above may be biased due to possible endogeneity. Lagged trading volume may be correlated with the error term of the regression in which the dependent variable is current volume due to the combination of two factors: positive autocorrelations of the fake trading measures, and a positive contemporaneous relation between fake trading measures and volume. To mitigate the endogeneity concern, we use an instrumental variable approach. In particular, we perform a two-stage estimation. In the first stage, reported in columns 3, 7, and 11 for the three fake trading measures, we regress lagged trading volume on a lagged fake trading measure and lagged price of Bitcoin (averaged over the course of the month) and its square.[38] Lagged Bitcoin price is unlikely to be correlated with lagged measure of fake trading at the exchange-currency pair level, i.e. it likely satisfies the exclusion restriction. Lagged volume is increasing in lagged Bitcoin price – the relation being concave, as evident from the negative coefficient on the squared Bitcoin price – suggesting that Bitcoin price satisfies the relevance restriction. Consistent with the results in the first two columns, lagged volume is positively related to lagged fake trading.

[38]The results are similar when the squared term is omitted from the regressions.

In the second stage, reported in columns 4, 8, and 12, we regress current trading volume on current and lagged volume-based fake trading measure, while replacing lagged volume by its fitted value from the first-stage regression. There are two interesting findings. First, the effect of past trading volume on current volume is half the size of that in the OLS regression in column 2. Second, the negative effect of lagged fake trading on current volume decreases substantially. Controlling for the instrument for past trading volume, the negative effect of lagged fake trading on current volume is statistically significant in only one specification (in column 12), potentially overturning the conclusion that market participants can partially see through volume inflation by exchanges.

Overall, the takeaway from Table 8 is that fake trading is, to a certain extent, an effective way for exchanges to affect future legitimate trading volume. This finding supports the conjecture that exchanges' use of volume-inflating strategies may generate added (legitimate) trading activity.

6.2 Real short-term and long-term effects of fake trading on operating performance

While the results in Table 8 indicate that fake trading may be effective in generating future trading volume, there is a possibility that the relation between past fake trading and current volume is due to inability of our fake trading measures to perfectly control for current volume inflation. To address this concern, in Table 9 we examine the effects of past fake trading on an exchange on alternative, non-volume-based outcomes, which are not subject to mechanical relation between volume inflation and reported trading volume.

[Insert Table 9 here]

In Panel A of Table 9, we estimate regressions, at the exchange-month level, of two measures of an exchange's performance – its web popularity and its estimated revenue from trading commissions – on measures of fake trading over the previous three months. Web popularity is measured as one minus the ratio of exchange's Alexa rank and the highest Alexa rank across exchanges (where the higher the rank the least popular the exchange). The relations between all three measures of fake trading on an exchange over the past three months and its current web popularity are positive, statistically significant, and economically large: A one-standard-deviation increase in lagged measures of fake trading is associated with a 1.5-1.9

standard-deviation increase in the exchange's web popularity. An interpretation is that investors do pay attention to rankings of exchanges, which are based in large part on reported trading volume.

To estimate an exchange's revenue from trading commissions in a given month, we adopt the following procedure. We begin by assuming that all the trading volume on an exchange is legitimate, and that the exchange is compensated for the entire volume. The compensation takes the form of trading commission, which we assume to be equal to 0.1%. This is a conservative estimate. Crypto exchanges' trading commissions typically range from 0.1% to 5%, depending on order size.[39] Since not all reported volume is legitimate, we attempt to estimate the proportion of reported volume that is fake and for which the exchange does not receive trading commissions. In particular, we ask the question: What would the volume on the exchange be if volume inflation did not take place? To answer this question, we perform the following exercise. First, using the estimates of the coefficients in the fourth column in Table 8, we compute the fitted value of volume for each exchange-currency pair-month.[40] Second, we aggregate these fitted values to obtain the predicted value of overall volume on an exchange in a given month. Third, we replace the value of the fake trading measure of a given exchange-currency pair-month by the lowest value of that measure on any exchange for the same currency pair, and compute hypothetical exchange-pair trading volume without volume inflation. Under the assumption that the exchange with the lowest fake trading measure does not inflate trading volume, this calculation results in an estimate of what the trading volume in a given currency pair on a given exchange would have been had the exchange not engaged in trading volume inflation. Fourth, we aggregate hypothetical trading volumes in all currency pairs on a given exchange to compute the overall hypothetical trading volume on the exchange in a given month had it not inflated volume in any of the currency pairs listed on it. Fifth, we compute the ratio of hypothetical aggregate fitted volume to true aggregate fitted volume on an exchange in a given month to obtain the estimated proportion of (il)legitimate trading. Our estimates indicate that the mean proportion of fake trading on an exchange is 19%, and the maximum is 87%. These are downward-biased estimates of the extent of fake trading. The reason is the assumption that the exchange with the lowest level of fake trading measure does not inflate its trading volume, which is likely an overly optimistic view. Finally, we multiply the product of reported volume and assumed trading commission by the estimated proportion of legitimate trading to

[39]According to https://www.Bitdegree.org/crypto, the cheapest exchanges and their respective fees are: Binance - up to 0.1%, Kraken - up to 0.26%, Cex - up to 0.25%, Bittrex - fixed at 0.25%, Coinbase - from 1.49% to 3.99%, Bitstamp - From 0.05% to 5%, and Poloniex - up to 0.125%.

[40]The results are similar when we use estimates from other regression specifications, such as those in columns 8 and 12 in Table 8.

arrive at an estimate of overall trading commissions on an exchange in a given month.

The effect of fake trading over the past three months on estimated exchange revenue is positive and significant both statistically and economically. A one-standard-deviation increase in lagged measures of fake trading is associated with 0.4-0.5 standard-deviation increase in estimated revenue of the exchange. The combination of this result with the positive relation between lagged fake trading and web popularity suggests that fake trading is effective in the short and medium term. These findings complement Cong et al. (2020), who report that wash trading has a positive impact on an exchange's rank on aggregator websites, such as `www.CoinMarketCap.com`.

In Panel B, we estimate longer-term effects of fake trading on non-volume-based measures of exchange performance. In particular, we measure past fake trading over a period of 12 months. The findings in Panel B are very different from those in Panel A. Both web popularity of an exchange and its estimated revenue from trading commissions are significantly negatively associated with all three measures of fake trading over the past year. In particular, an increase in a measure of past fake trading by one standard deviation is associated with 2.7-3.4 standard-deviation reduction in web popularity and with 0.2-0.3 standard-deviation reduction in estimated trading commissions.

The finding that fake trading has positive short-term effects – on both legitimate trading volume (in Table 8) and non-volume-based performance indicators (in Panel A of Table 9), and negative long-term effects (in Panel B of Table 9) suggests that volume inflation has short-term benefits and long-term costs. This result is consistent with exchanges choosing the degree of volume inflation while trading off near-term increase in operating performance against longer-term adverse reputation effects of fake trading.

7 Conclusion

We examine the effects of competition among crypto exchanges on their incentives to engage in an opportunistic strategy of inflating (faking) trading volume. The benefit of this strategy is that an exchange with (seemingly) large trading volume is attractive to potential traders, as larger volume is likely to be associated with lower direct and indirect trading costs. The cost of volume inflation is reputation damages, which occur once investors realize that not all reported trading volume is legitimate. Theoretical effects of competition among exchanges on this tradeoff are ambiguous. On the one hand, reputation is more valuable when future rents are higher, i.e. when there is less competition. On the other hand, competition creates outside options for consumers, leading to enhanced importance of reputation.

Our main findings consistently point to the positive relation between competition and fake trading. Measures of static competition, such as the number of exchanges trading in a given currency pair and the Herfindahl index of exchanges, indicate that trading volume inflation is increasing in competition. An analysis of effects of entry and exit by various types of competitor exchanges – general, geographical, and operational – in particular currency pairs suggests that fake trading intensifies upon competitor exchange entries and declines upon competitor exchange exits, especially when the ex-ante landscape is highly competitive.

Our analysis of the effectiveness of fake trading, i.e. its ability to generate short-term rents, shows that exchanges are generally successful in misleading traders in the short run. Fake trading raises future trading volume, exchange's web popularity, and estimated trading commissions over one-to-three months horizons. However, the strategy of inflating trading volume has its long-term costs. The effects of fake trading on longer-run operating performance is negative: Both the web popularity and estimated trading commissions are significantly negatively related to fake trading over the twelve-month horizon.

While our sample is large – it includes 41 crypto exchanges and over 1,700 currency pairs, it is currently biased towards larger, more important currencies and largest exchanges. To the extent that fake trading is more prevalent in smaller, less reputable exchanges – a result that we observe in the data – our estimates of fake trading may be biased downwards, and the effects of competition on fake trading may be understated. We plan to expand the sample to include all exchanges and currency pairs available on www.Kaiko.com, as well as to update the sample period by including data post September 2019.

We use the crypto exchange industry as a convenient laboratory for examining the general effects of competition on firms' product quality choices that extend beyond this industry. In addition to the general implications, our analysis also has particular implications for future regulation of crypto exchanges. Our finding of pervasive volume inflation complements nascent empirical literature on crypto exchanges (e.g., Amiram et al. (2020) and Cong et al. (2020)) in illustrating the consequences of generally lax regulation of and low compliance by exchanges, and in highlighting the importance of regulation of the crypto exchange industry, which is one of the cornerstones of the crypto market as a whole.

References

Abbott, L. (1955). Quality and competition: an essay in economic theory. Columbia University Press.

Ajinkya, B. and Jain, P. (1989). A test of dividend irrelevance using volume reactions to a change in dividend policy. Journal of Accounting and Economics, 4(11):331–359.

Allen, F. (1984). Reputation and product quality. RAND Journal of Economics, 15(3):311–327.

Aloosh, A. and Li, J. (2020). Direct evidence of bitcoin wash trading. George Mason University working paper.

Amiram, D., Bozanic, Z., and Rouen, E. (2015). Financial statement errors: Evidence from the distributional properties of financial statement numbers. Review of Accounting Studies, 20(4):1540–1593.

Amiram, D., Jørgensen, B. N., and Rabetti, D. (2020). Coins for bombs: Increased transparency of the global financial system - evidence from terrorist attacks financing detection in blockchain-based currencies. Tel Aviv University woworking paper.

Bar-Isaac, H. and Shapiro, J. (2013). Ratings quality over the business cycle. Journal of Financial Economics, 108(1):62–78.

Becker, B. and Milbourn, T. (2011). How did increased competition affect credit ratings? Journal of Financial Economics, 101(3):493–514.

Benford, F. (1938). The law of anomalous numbers. Proceedings of the American Philosophical Society, 78(4):551–72.

Bennett, V. M., Pierce, L., Snyder, J. A., and Toffel, M. W. (2013). Customer-driven misconduct: How competition corrupts business practices. Management Science, 59(8):1725–1742.

Bolton, P., Freixas, X., and Shapiro, J. (2012). The credit ratings game. Journal of Finance, 67(1):85–111.

Chamberlin, E. (1933). Monopolistic competition. Journal Of Business Economics Research, pages 17–28.

Chemmanur, T. J. and Fulghieri, P. (1994). Investment bank reputation, information production, and financial intermediation. Journal of Finance, 49(1):57–79.

Cong, W., Li, X., Tang, K., and Yang, Y. (2020). Crypto wash trading. Cornell University working paper.

Conover, W. (1971). Practical nonparametric statistics. new york: John wiley & sons. pages 295–301.

Cooper, R. and Ross, T. W. (1984). Prices, product qualities and asymmetric information: The competitive case. Review of Economic Studies, 51(2):197–207.

Cumming, D., Johan, S., and Li, D. (2011). Exchange trading rules and stock market liquidity. Journal of Financial Economics, 99(3):651–671.

Dana, J. J. D. and Fong, Y. F. (2011). Product quality, reputation, and market structure. International Economic Review, 52(4):1059–1076.

Domberger, S. and Sherr, A. (1989). The impact of competition on pricing and quality of legal services. International Review of Law and Economics, 1(9):41–56.

Drake, P. D. and Nigrini, M. J. (2000). Computer assisted analytical procedures using benford's law. Journal of Accounting Education, 18:127–146.

Fusaro, T. and Hougan, M. (2019). Presentation to the us securities and exchange commission. (Bitwise Asset Management).

Gandal, N., Hamrick, J. T., Moore, T., and Oberman, T. (2018). Price manipulation in the bitcoin ecosystem. Journal of Monetary Economics, 95:86–96.

Gervais, S., Kaniel, R., and Mingelgrin, D. H. (2001). The high-volume return premium. Journal of Financial Economics, 56(3):877–919.

Griffin, J. M. and Shams, A. (2020). Is bitcoin really untethered? Journal of Finance, 75(4):1913–1964.

Grinblatt, M. and Keloharju, M. (2004). Tax-loss trading and wash sales. Journal of Financial Economics, 71(1):51–76.

Horner, J. (2002). Reputation and competition. American Economic Review, 92(3):644–663.

James, N. A., Kejariwal, A., and Matteson, D. S. (2016). Leveraging cloud data to mitigate user experience from 'breaking bad'. IEEE International Conference on Big Data.

James, N. A. and Matteson, D. S. (2015). ecp: An r package for nonparametric multiple change point analysis of multivariate data. Journal of Statistical Software, 62(7).

Karpoff, J. M., Lee, D. S., and Martin, G. S. (2008). The cost to firms of cooking the books. Journal of Financial and Quantitative Analysis, 43(3):581–611.

Klein, B. and Leffler, K. B. (1981). The role of market forces in assuring contractual performance. Journal of Political Economy, 89(4):615–641.

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. Management Science, 62(12):3412–3427.

Lyandres, E., Palazzo, D., and Rabetti, D. (2020). Ico success and post-ico performance. Tel Aviv University working paper.

Mailath, G. J. and Samuelson, L. (2001). Who wants a good reputation? Review of Economic Studies, 68(2):415–441.

Mathis, J., McAndrews, J., and Rochet, J. C. (2009). Rating the raters: are reputation concerns powerful enough to discipline rating agencies? Journal of Monetary Economics, 56(5):657–674.

Matsa, D. (2011). Competition and product quality in the supermarket industry. Quarterly Journal of Economics, 3(126):1539–1591.

Mazzeo, M. (2003). Competition and service quality in the us airline industry. Review of Industrial Organization, 4(22):275–296.

Michalski, T. and Stoltz, G. (2013). Financial statement errors: Evidence from the distributional properties of financial statement numbers. Review of Economics and Statistics, 2(95):591–616.

Pimbley, J. (2014). Benford's law and the risk of financial fraud. Risk Professional, 5:1–7.

Richardson, G., Sefcik, S., and Thompson, R. (1986). A test of dividend irrelevance using volume reactions to a change in dividend policy. Journal of Financial Economics, 2(17):313–333.

Schwenkler, G. and Zheng, H. (2019). The network of firms implied by the news. Santa Clara University working paper.

Schwenkler, G. and Zheng, H. (2020). Competition or contagion? evidence from cryptocurrency peers. Santa Clara University working paper.

Shams, A. (2020). The structure of cryptocurrency returns. Ohio State University working paper.

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. Quarterly Journal of Economics, 98(4):659–679.

Shleifer, A. (2004). Does competition destroy ethical behavior? American Economic Review, 2(94):414–418.

Figure 1. Evolution of crypto market capitalization. The plot shows the evolution of market capitalisation in billions of US Dollars for Bitcoin (BTH, black solid), Ether (ETH, dashed blue) and remaining coins and tokens (dotted red). Market capitalizations are aggregated monthly in the period from January 2016 to September 2019.

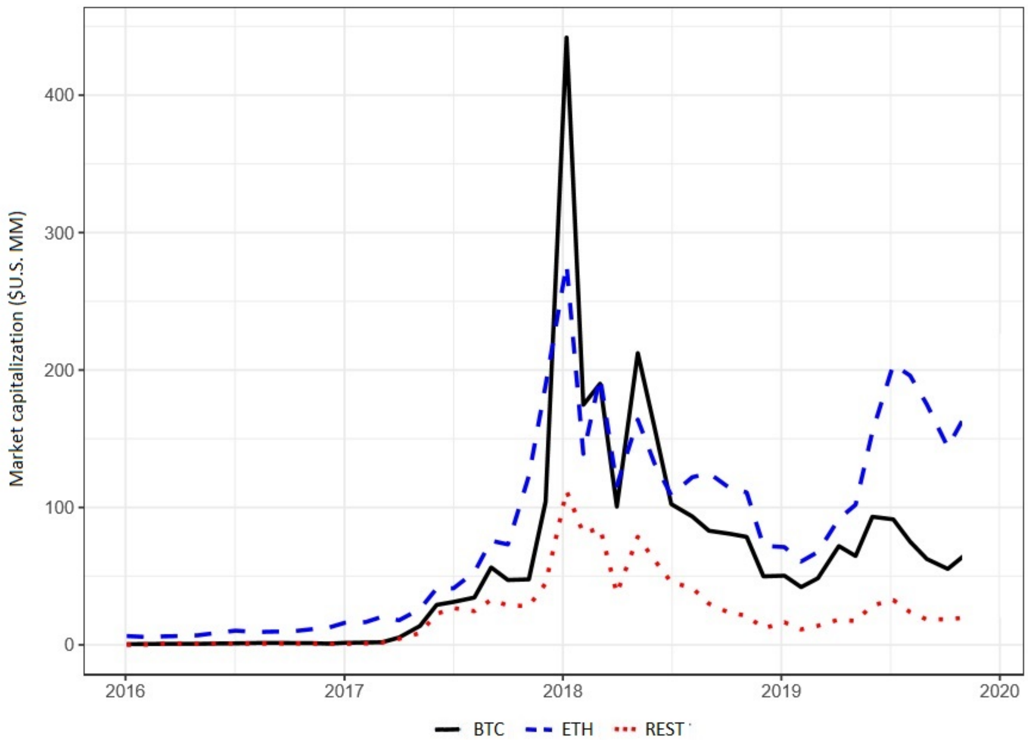


Figure 2. Dynamics of entry and exit of currency pairs. The plot depicts the number of entries (solid blue), the number of exits (dashed red) and entries net of exits (grey bars) of currency pairs on crypto exchanges. The numbers of entries and exists are computed at the currency pair-exchange level and are aggregated monthly in the period from January 2016 to September 2019.

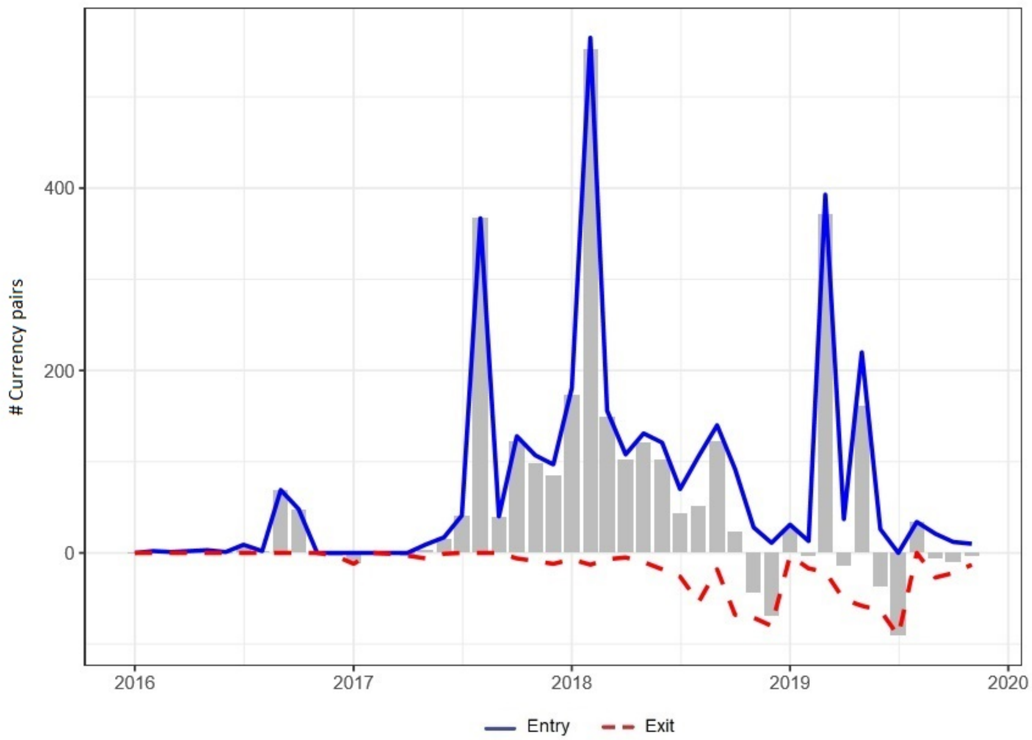
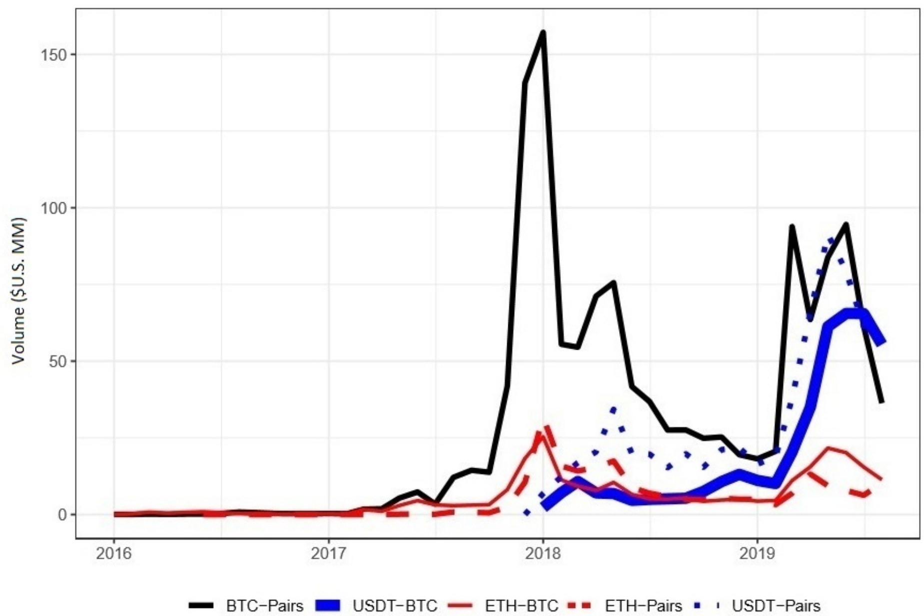


Figure 3. Evolution of trading volume and number of trades. The plot shows the evolution of reported trading volume (in Panel A) and number of trades (in Panel B) for ETH-BTC pair (solid red), USDT-BTC pair (solid blue), other BTC pairs (solid black), other ETH pairs (dashed red), and other USDT pairs (dotted blue). Trading volume is aggregated monthly in the period from January 2016 to September 2019 and reported in billions of $U.S. Number of trades is aggregated monthly and reported in millions.

Panel A: Volume



Panel B: # Trades

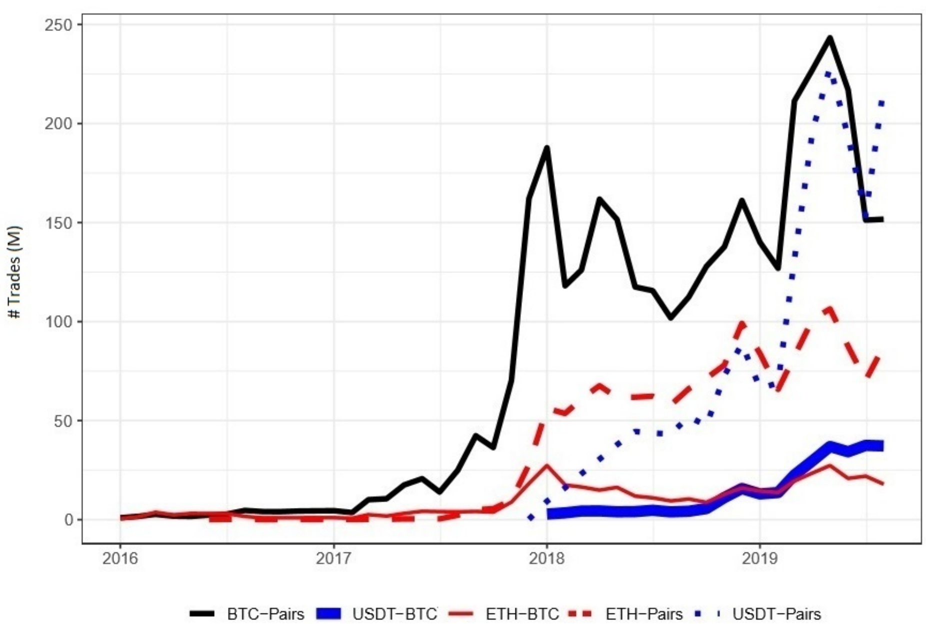


Figure 4. Evolution of exchange market concentration. The plot shows the evolution of Herfindahl index across exchanges based on number of currency pairs (solid black), trading volume (dashed blue), and number of trades (dotted red). Currency-pair-based Herfindahl index is computed as the ratio of the sum of squared number of currency pairs traded on each exchange to the squared sum of all numbers of currency pairs traded on all exchanges. Trading-volume-based and number-of-trades-based Herfindahl indices are computed similarly. The grey bars show the total number of distinct currency pairs. The measures are computed monthly during the period from January 2016 to September 2019.

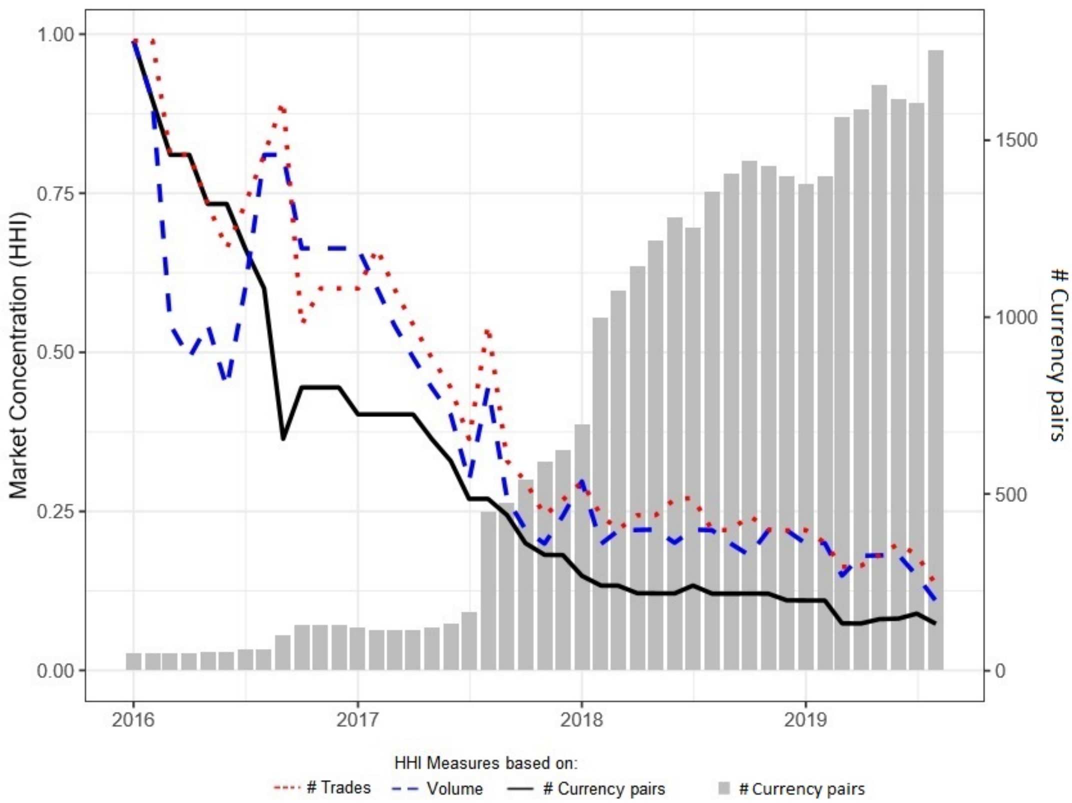


Figure 5. Examples of trading volume and number of trades series. The plot shows trading volume (in upper figures of Panels A, B, and C) and number of trades (in the lower figures of Panels A, B, and C), aggregated into 4,032-4,464 ten-minute bins. Panel A displays data for OMG-BTC pair on Okex (left panels) and Binance (right panels) in January 2019. Panel B displays data for ETH-BTC pair on ZB (left panels) and Binance (right panels) in March 2019. Panel C displays data for TNB-BTC pair on Huobi (left panels) and Binance (right panels) in February 2019.

Panel A

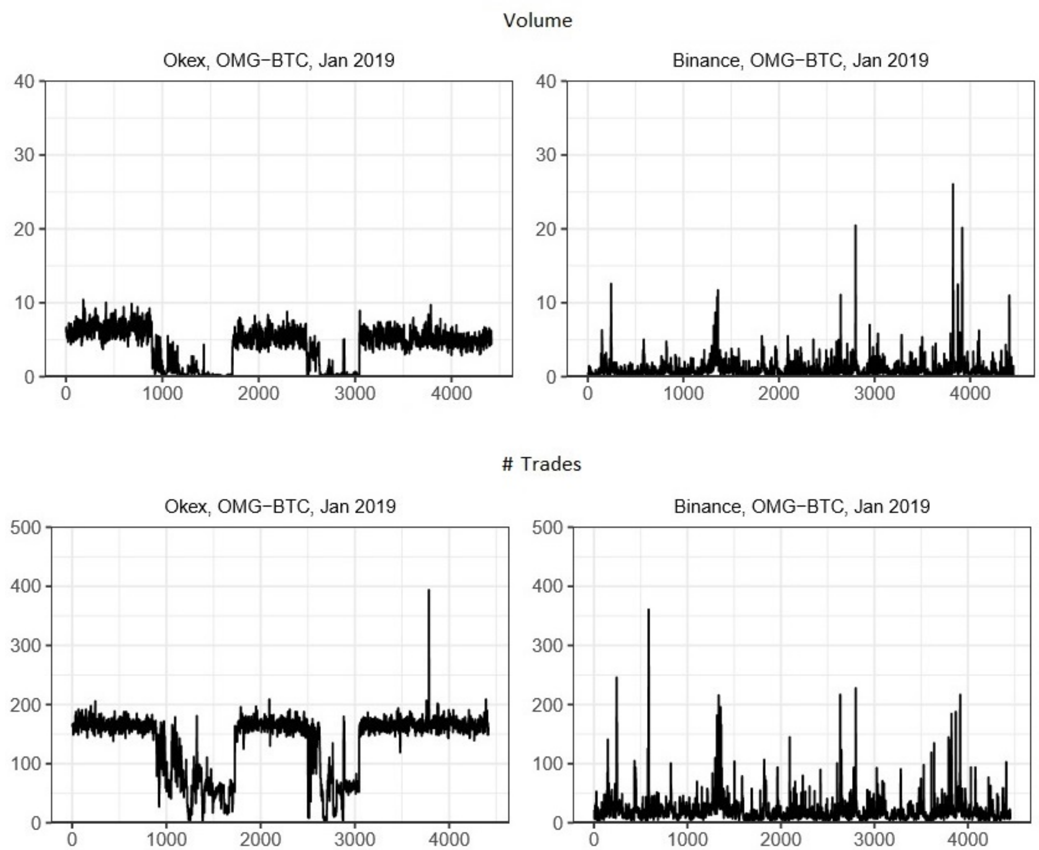


Figure 5. Examples of trading volume and number of trades series – continued

Panel B

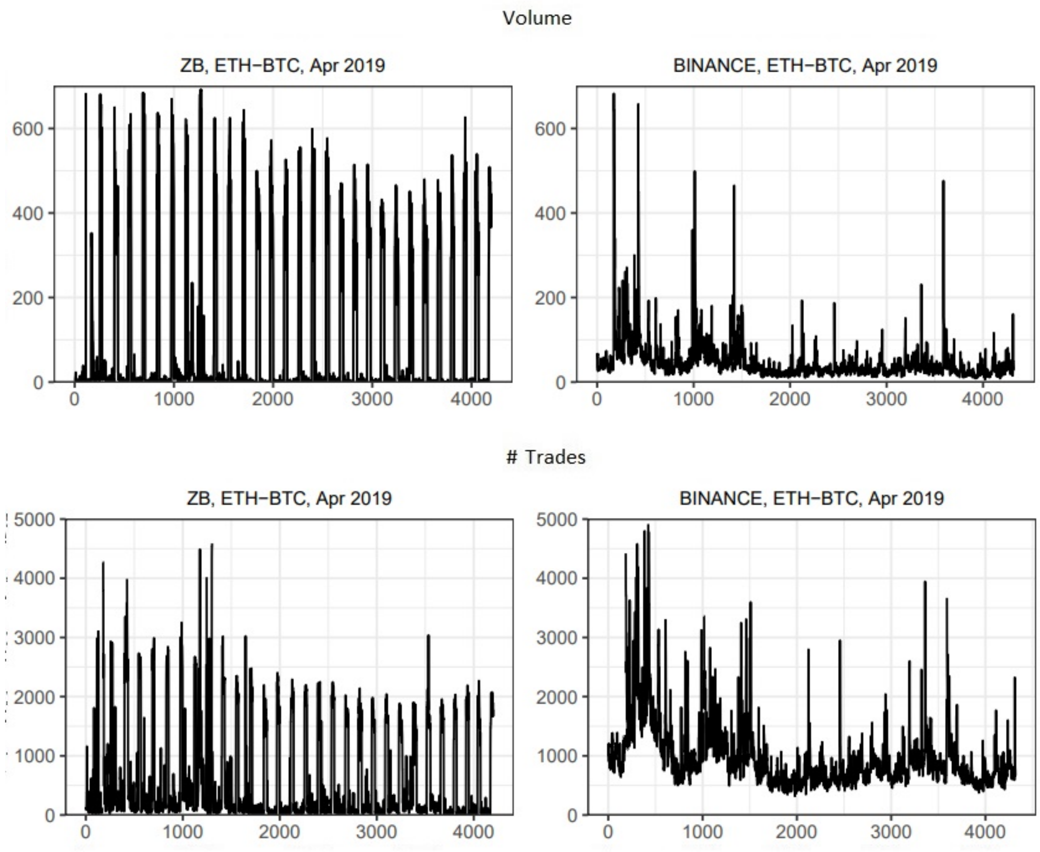


Figure 5. Examples of trading volume and number of trades series – continued

Panel C

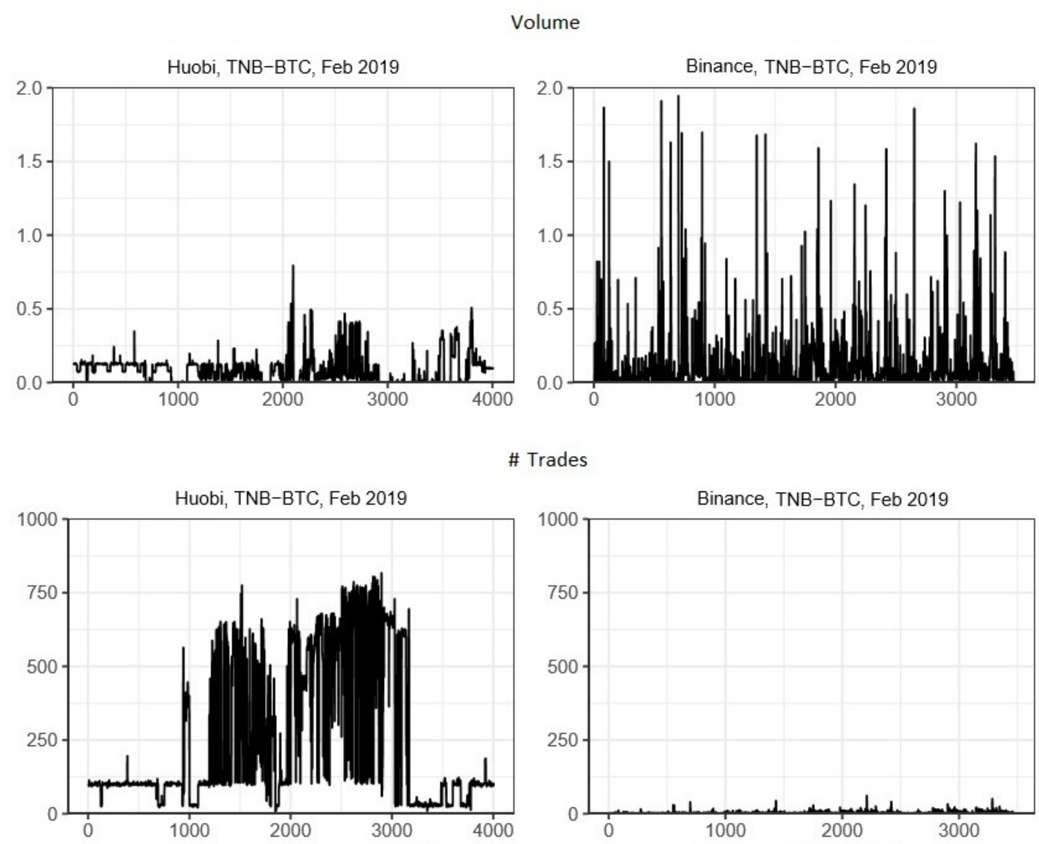


Figure 6. Examples of deviations from Benford's Law. The plot shows empirical frequencies of leading digits of trading volume (in Panel A, blue blue curve) and number of trades (in Panel B, blue solid line) aggregated into ten-minute bins within a month. Dotted red curve represents frequencies of leading digits under Benford's Law. The frequencies are computed for ETH-BTC pair in March 2019 on four exchanges: Binance, ZB, Okex, and Bibox.

Panel A: Trading volume

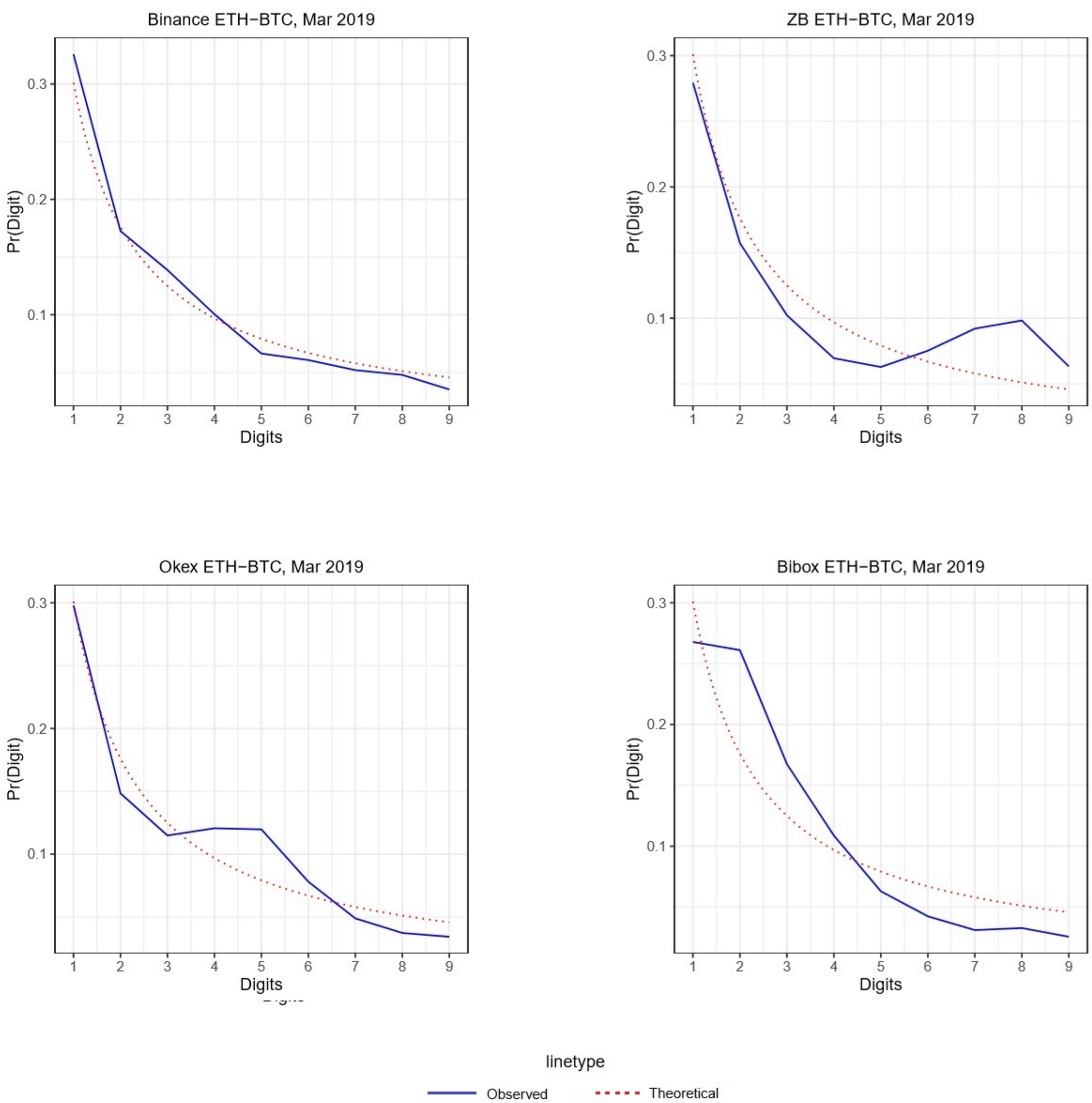


Figure 6. Examples of deviations from Benford's Law – continued

Panel B: # Trades

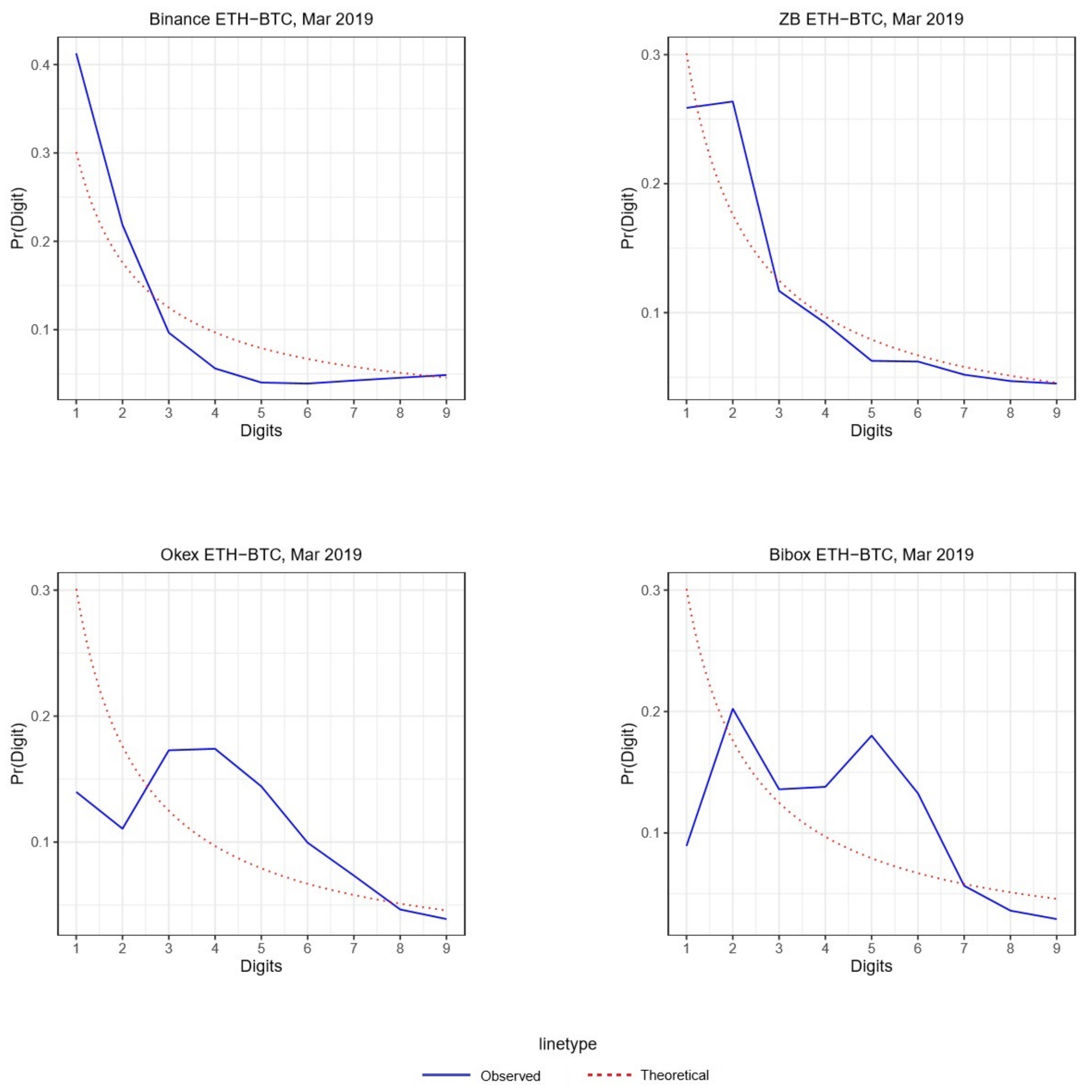


Figure 7. Examples of deviations from log-normal distribution. The plot shows empirical cumulative density function (solid blue curve) of the natural logarithm of trading volume (in Panel A) and number of trades (Panel B) aggregated into ten-minute bins within a month, and c.d.f. of normal distribution with the same mean and variance (dotted red curve). The c.d.f.'s are computed for ETH-BTC pair in March 2019 on four exchanges: Binance, ZB, Okex, and Bibox.

Panel A: Trading volume

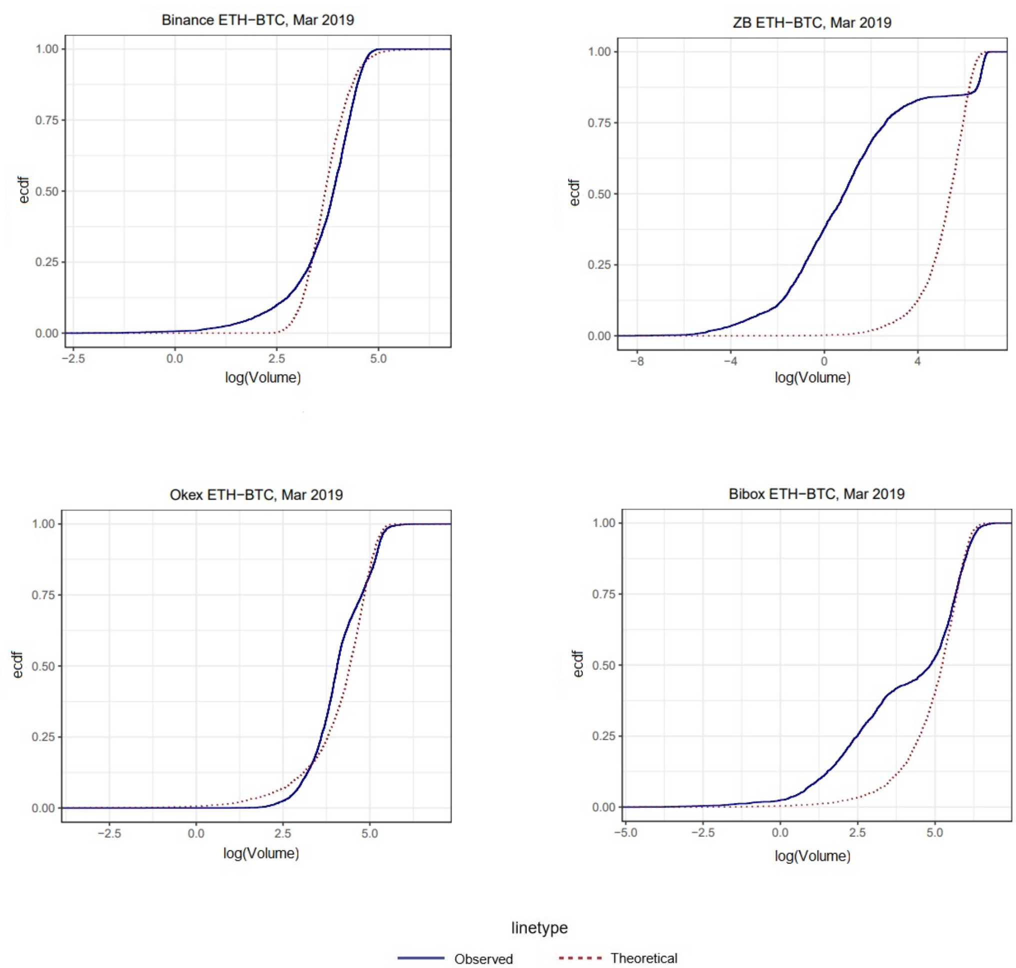


Figure 7. Examples of deviations from log-normal distribution – continued

Panel B: # Trades

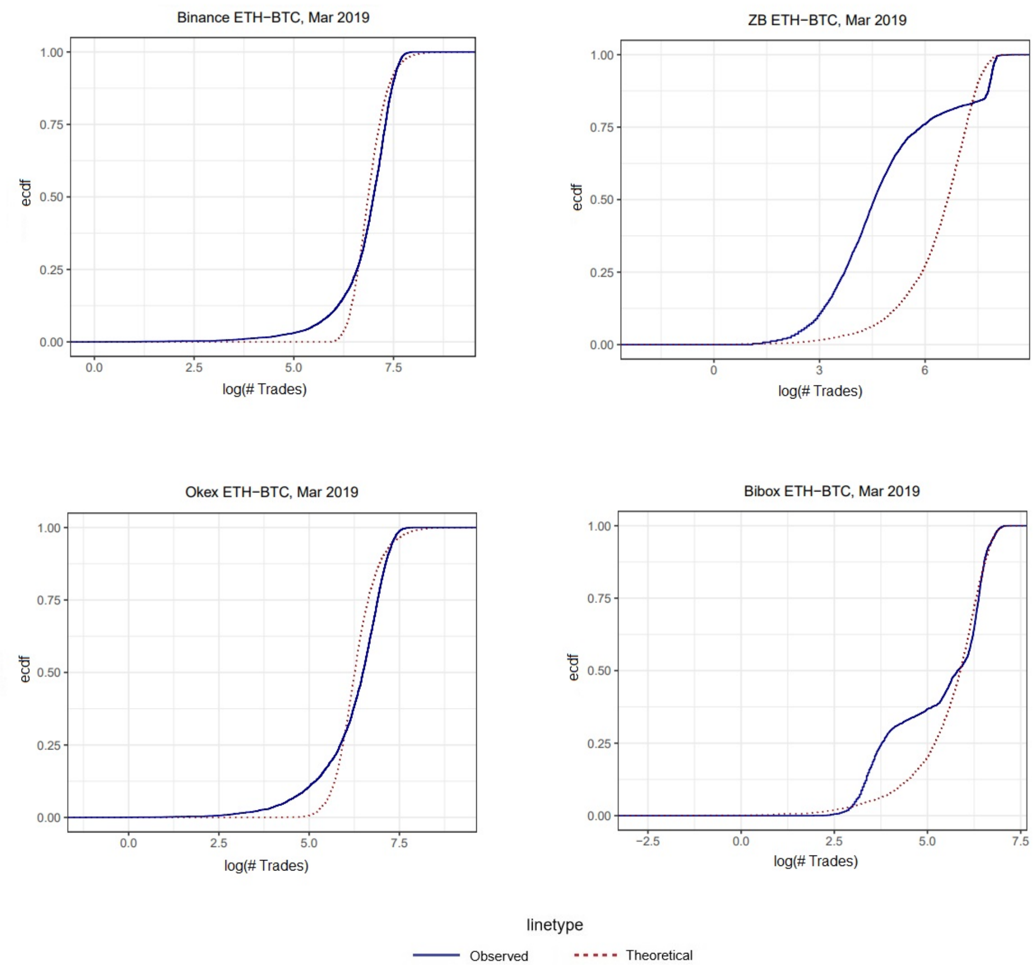
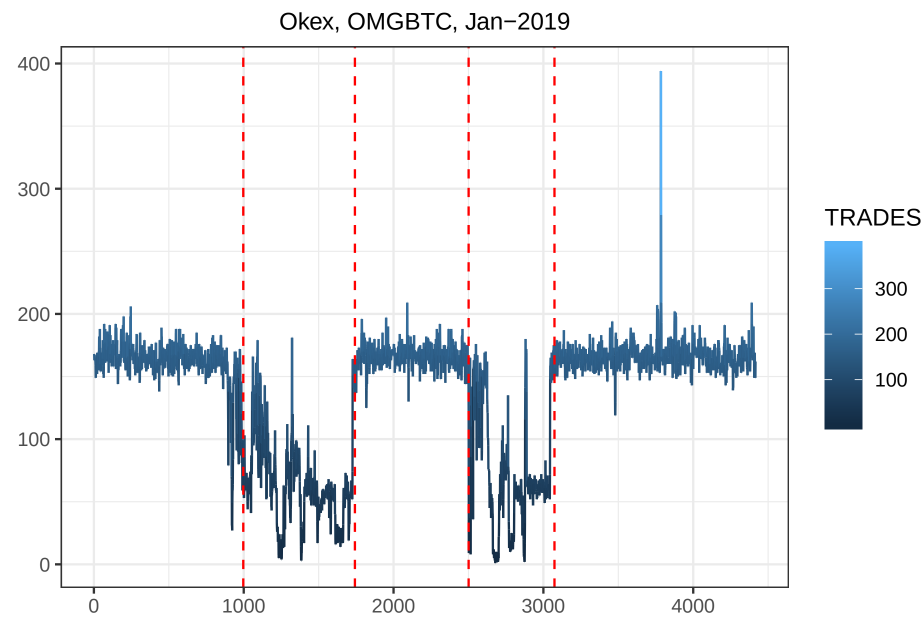


Figure 8. Examples of application of EDM measure. The plot in Panel A shows structural breaks (dashed red lines) in the number of trades series, aggregated into ten-minutes bins within a month, on Okex exchange in OMG-BTC pair in January 2019. The plot in Panel B depicts the number of trades series, aggregated into ten-minutes bins within a month, on Binance exchange in ETH-BTC pair in May 2019, in which EDM algorithm did not detect any structural breaks.

Panel A



Panel B

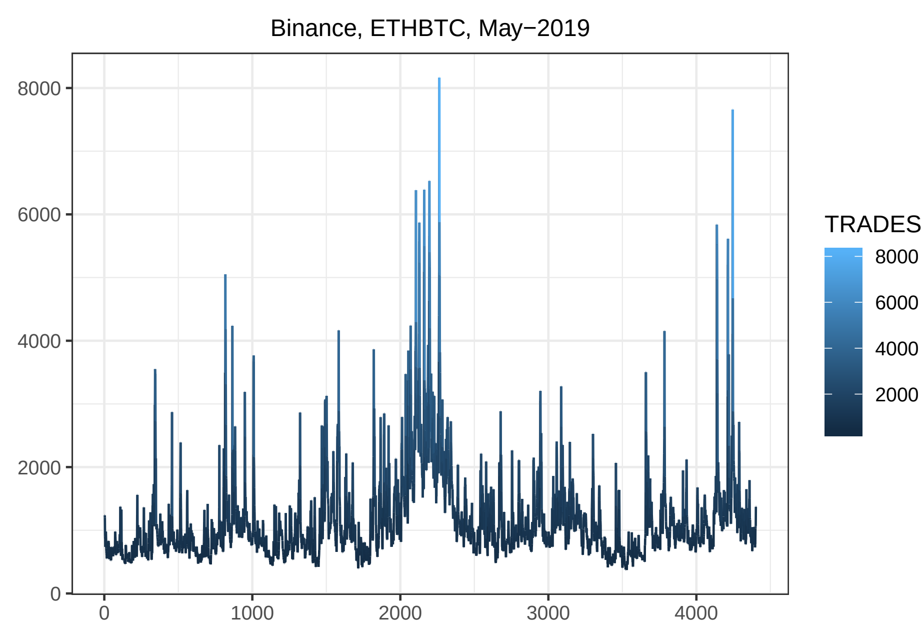


Figure 9. Principle components - Scree plots. The three plots show the percentages of variation in the fake trading measures explained by the first three principal components. In Panel A, the principal component, PC1: Volume, is the principal component of volume-based fake trading measures (MAD: Volume, KS: Volume, and EDM: Volume). In Panel B, the principal component, PC1: Trades is the principal component of number-of-trades-based fake trading measures (MAD: # Trades, KS: # Trades, and EDM: # Trades). In Panel C, the principal component, PC1: Both, is the principal component of both volume-based and number-of-trades-based fake trading measures.

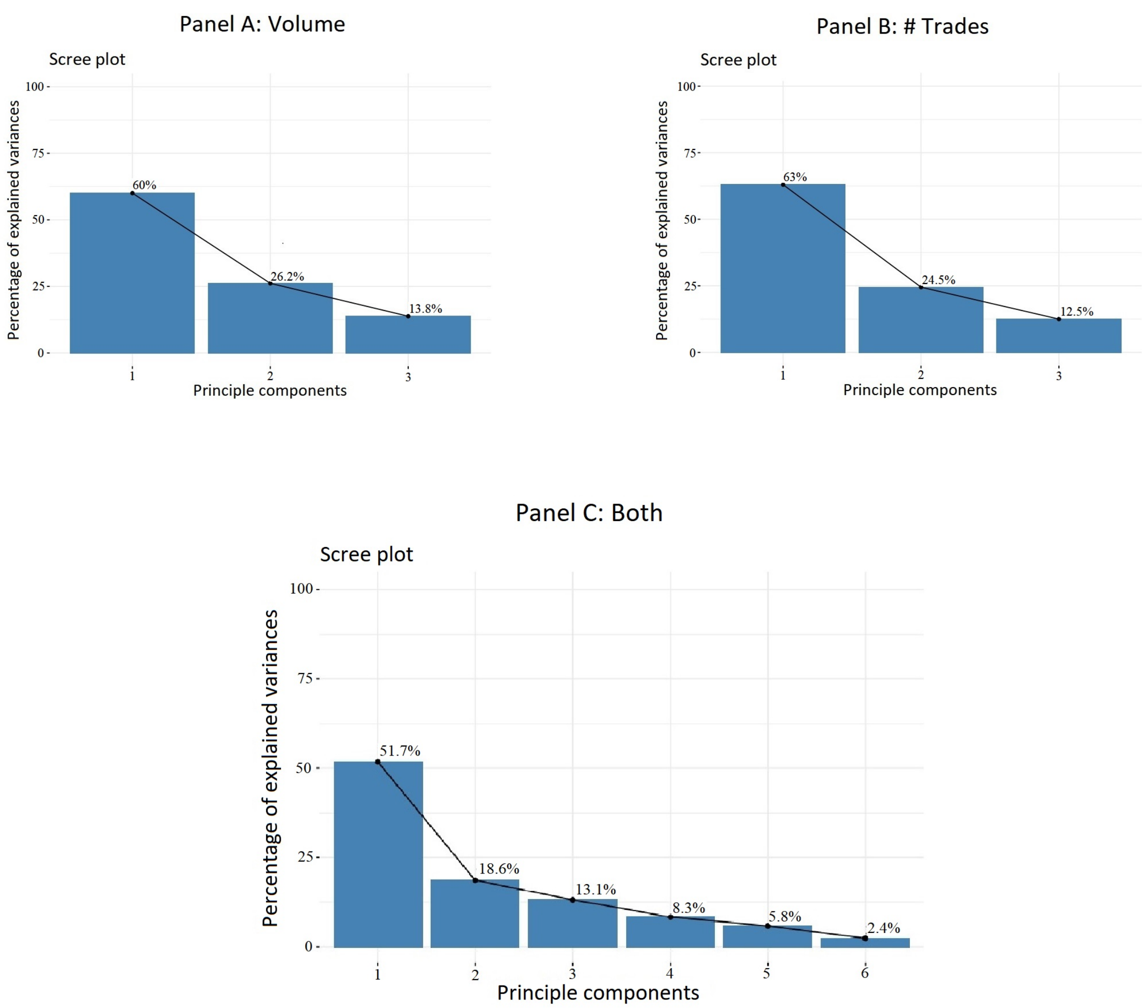


Figure 10. Principle components - Biplots. The three plots depict the proportion of variation in the first principal component explained by measures of fake trading. The length of each vector (squared cosine) represents the respective variable in the first two principal components. The horizontal and vertical projections of a vector represent the variable in the first and second principal component respectively. In Panel A, the principal component, PC1: Volume, is the principal component of volume-based fake trading measures (MAD: Volume, KS: Volume, and EDM: Volume). In Panel B, the principal component, PC1: Trades is the principal component of number-of-trades-based fake trading measures (MAD: # Trades, KS: # Trades, and EDM: # Trades). In Panel C, the principal component, PC1: Both, is the principal component of both volume-based and number-of-trades-based fake trading measures.

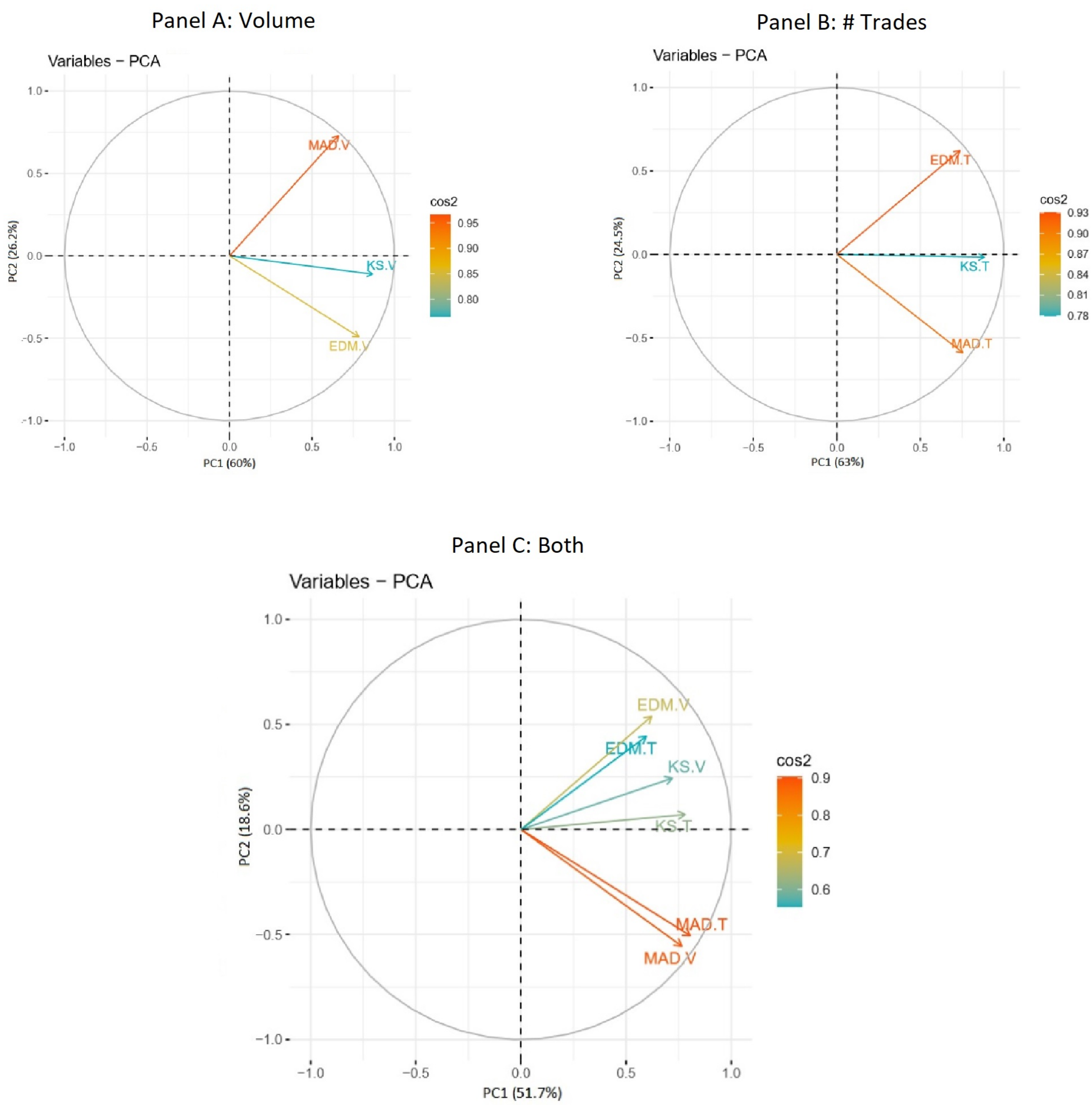


Figure 11. (A) Mean fake trading measures over time. The plot shows equally-weighed (across exchange-currency pair-month) means of the three principle-component-based fake trading measures – PC1: Volume (blue squares), PC1: # Trades (orange triangles), and PC1: Both (green circles) – for each quarter in our sample period between 2013-Q2 to 2019-Q3. The lines around point estimates indicate confidence intervals.

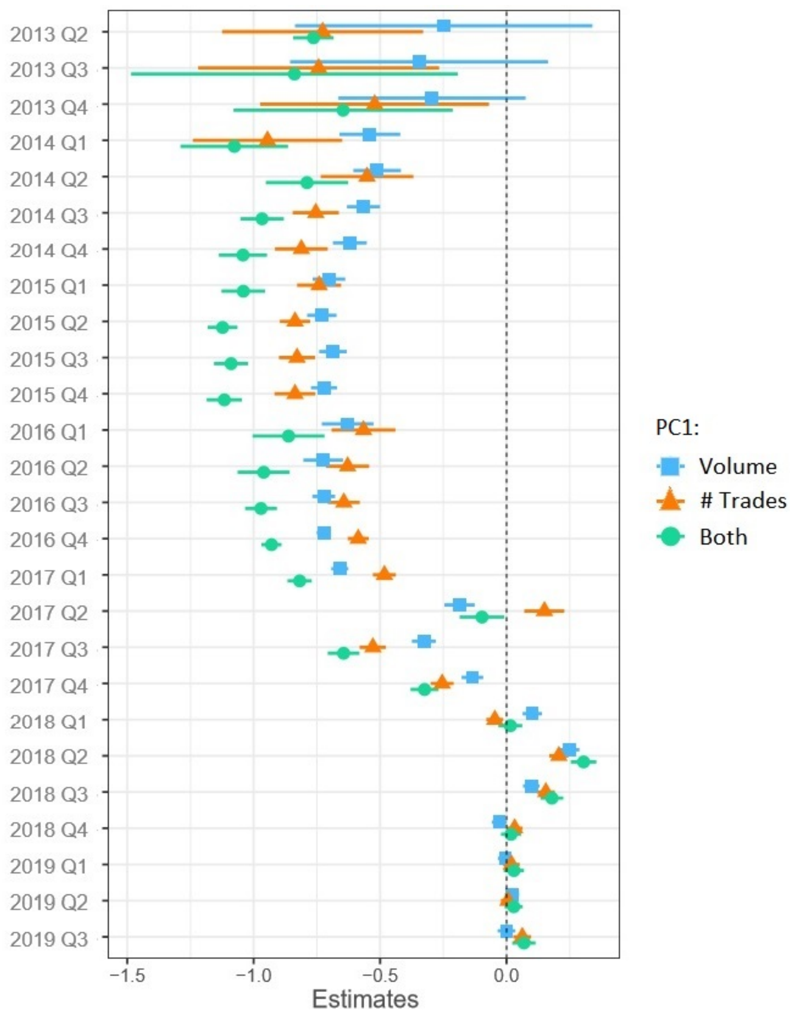


Figure 11. (B) Mean fake trading measures by subsets of currency pairs over time. The plot shows equally-weighed (across exchange-currency pair-month) means of combined principle-component-based fake trading measure – PC1: Both – for each quarter in our sample period between 2013-Q2 to 2019-Q3, and for each of the five groups of currency pairs: ETH-BTC (green squares), USDT-BTC (dark blue circles), other BTC pairs (light blue squares), other ETH pairs (orange crosses), and other USDT pairs (pink triangles). The lines around point estimates indicate confidence intervals.

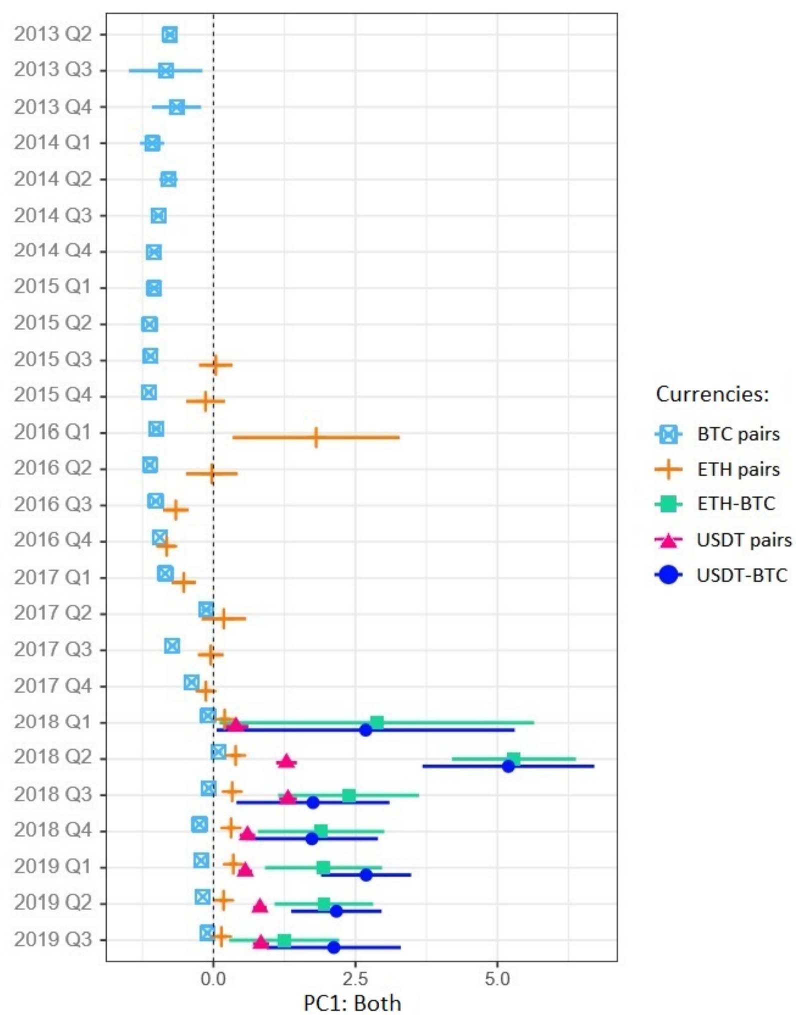


Figure 11. (C) Mean fake trading measures by exchange. The plot shows equally-weighed (across currency pair-month) means of the three principle-component-based fake trading measures – PC1: Volume (blue squares), PC1: # Trades (orange triangles), and PC1: Both (green circles) – for each exchange in our sample period between 2013-Q2 to 2019-Q3. The lines around point estimates indicate confidence intervals.

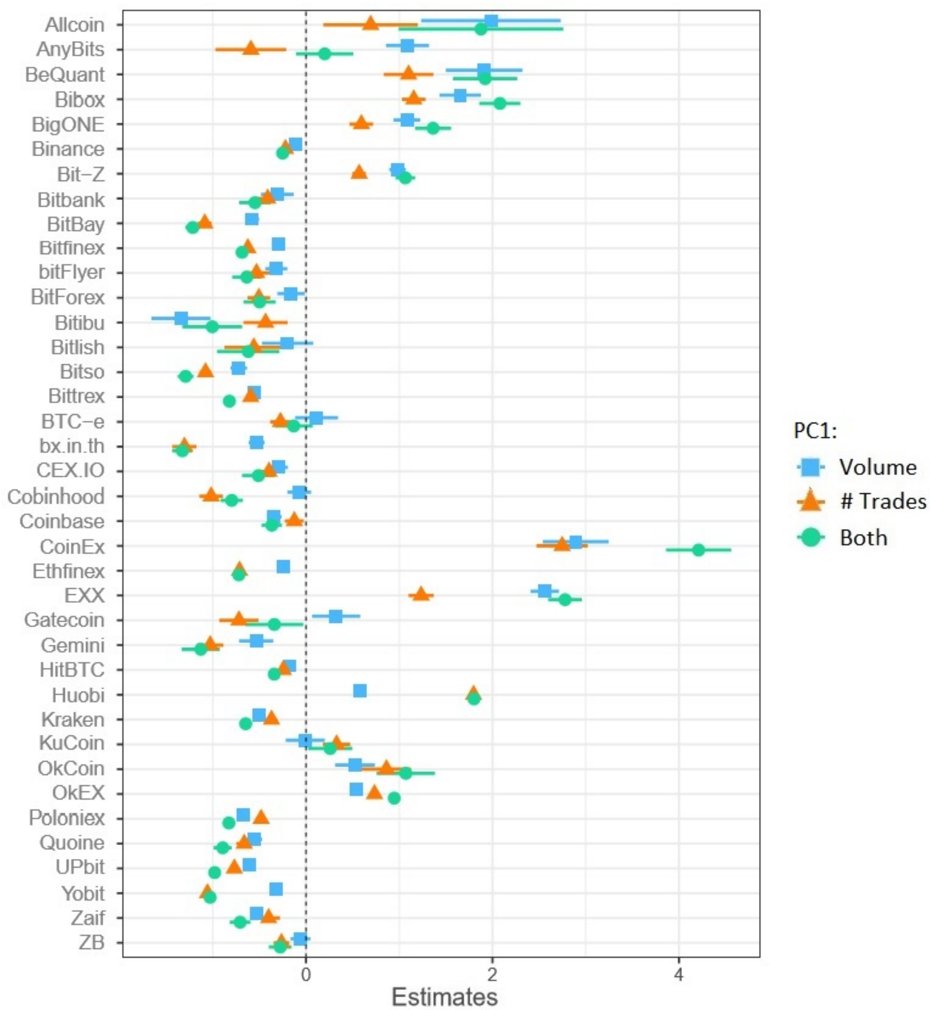


Table 1. Summary statistics. The table reports summary statistics for market-level (in Panel A), currency-pair-level (in Panel B), exchange-level (in Panel C), and exchange-currency-pair level (in Panel D) of the variables used in the empirical analysis. The sample period is May 2013 – September 2019. All variables are defined in Table A.1 in the Appendix.

Panel A: Market

	Min	Max	Median	Mean	SD	Obs.
Market cap: CoinMarketCap (CMC)	0.18	1,046.85	12.13	184.73	342.97	76
Market cap: Kaiko	0.01	1,003.16	10.05	178.58	287.06	76
Market cap Kaiko / Market cap CMC	0.05	0.96	0.82	0.87	0.21	76
Currencies: All	1.00	967.00	54.50	262.41	326.03	76
Currencies: Tokens	0.00	626.00	21.50	148.91	205.22	76
Currencies: Coins	1.00	348.00	33.00	113.76	123.32	76
Currencies: Entry	0.00	380.00	2.50	33.03	68.81	76
Currencies: Exit	0.00	80.00	0.00	7.62	16.25	76
Currency pairs: All	1.00	1,753.00	56.50	428.01	590.13	76
Currency pairs: Tokens	0.00	1,221.00	22.50	274.32	405.77	76
Currency pairs: Coins	1.00	532.00	34.00	153.70	185.96	76
Currency pairs: Entry	0.00	593.00	2.50	53.72	117.99	76
Currency pairs: Exit	0.00	91.00	0.00	9.67	20.98	76
Volume: $U.S. (MM)	0.00	267.59	0.54	37.87	68.66	76
Volume: BTC (M)	0.00	43.14	0.76	5.63	9.29	76
# Trades (M)	0.00	642.66	5.06	99.4	166.4	76
Exchanges	1.00	41.00	9.50	10.86	10.29	76
Exchanges: Entry	0.00	14.00	0.00	0.53	1.73	76
Exchanges: Exit	0.00	1.00	0.00	0.05	0.22	76
HHI Exchanges: # Currencies	0.07	1.00	0.12	0.14	0.13	76
HHI Exchanges: Volume	0.11	1.00	0.20	0.22	0.11	76
HHI Exchanges: # Trades	0.13	1.00	0.22	0.24	0.12	76

Panel B: Currency pairs

	Min	Max	Median	Mean	SD	Obs.
Listed on # exchanges	1.00	19.00	1.81	2.46	0.93	32,529
HHI: Currency pair across exchanges: Volume	0.11	1.00	0.20	0.23	0.13	32,529
HHI: Currency pair across exchanges: # Trades	0.13	1.00	0.22	0.26	0.14	32,529
Age of listing on any exchange	0.00	75.00	9.00	11.63	10.56	32,529
Age of listing on a given exchange	0.00	61.00	8.00	8.92	6.94	32,529
Time to listing	0.00	36.00	4.22	5.87	4.55	32,529
Token	0.00	1.00	0.51	0.64	0.48	32,529

Table 1. Summary statistics – continued

Panel C: Exchanges

	Min	Max	Median	Mean	SD	Obs.
Age	1.00	8.00	5.00	4.95	1.69	807
Market share: Volume	0.00	1.00	0.006	0.094	0.20	807
Market share: # trades	0.00	1.00	0.007	0.094	0.19	807
Market share: Currency pairs	0.00	1.00	0.006	0.094	0.20	807
Currency pairs	1.00	581.00	12.00	77.67	122.14	807
Currency pairs: Entry	0.00	419.00	0.00	5.99	27.09	807
Currency pairs: Exit	0.00	62.00	0.00	0.95	4.16	807
AML	0.00	1.00	1.00	0.59	0.49	807
KYC	0.00	1.00	1.00	0.62	0.49	807
Crypto-friendly location	0.00	1.00	0.00	0.46	0.50	807
Bad news	0.00	1.00	0.00	0.25	0.43	807
Multiplatform	0.00	1.00	0.00	0.12	0.32	807
Alexa (K)	0.23	735.99	9.47	41.75	98.02	342
Reddit	0.00	2,228.00	0.00	37.85	156.51	418
Twitter	0.00	21,002.00	0.00	178.45	1,133.73	442
Github	0.00	695.00	13.00	103.36	159.94	237
Africa	0.00	1.00	0.00	0.02	0.15	807
Asia	0.00	1.00	0.00	0.25	0.43	807
China	0.00	1.00	0.00	0.14	0.35	807
Central and South America	0.00	1.00	0.00	0.12	0.33	807
Eastern Europe	0.00	1.00	0.00	0.09	0.28	807
Western Europe	0.00	1.00	0.00	0.12	0.33	807
Europe: Islands	0.00	1.00	0.00	0.18	0.38	807
North America	0.00	1.00	0.00	0.08	0.26	807

Panel D: Exchange-Currency pairs

	Min	Max	Median	Mean	SD	Obs.
Volume: $U.S. (M)	0.00	49,374.67	0.78	45.92	413.00	62,676
Volume: BTC (K)	0.00	27,260	6.94	9.33	7.24	62,676
# Trades (K)	1.00	25,593.80	12.32	120.53	447.24	62,676
HHI Currency pairs within exchange: Volume	0.02	1.00	0.50	0.51	0.36	62,676
HHI Currency pairs within exchange: # Trades	0.02	1.00	0.50	0.51	0.36	62,676

Table 2. Fake trading measures and principal components. Panel A reports summary statistics of three volume-based and three number-of-trades-based measures of fake trading (MAD, KS, and EDM). All measures are described in Table A.1 in the Appendix. Panel B reports summary statistics of the first principal components of the fake trading measures. Fake trading measures and principal components are at the exchange-currency pair-month level.

Panel A: Quality measures

	Min	Max	Median	Mean	SD	Obs.
MAD: Volume	0.38	463.80	12.43	31.63	42.64	55,242
MAD: # Trades	0.34	241.99	7.56	17.57	26.82	55,242
KS: Volume	0.00	0.53	0.33	0.31	0.08	55,228
KS: # Trades	0.00	0.56	0.29	0.27	0.08	55,228
EDM: Volume	0.00	1239.00	8.00	18.73	43.60	57,004
EDM: # Trades	0.00	885.00	3.00	11.47	33.93	57,004

Panel B: Principle components

	Min	Max	Median	Mean	SD	Obs.
PC1: Volume	-2.30	15.63	-0.44	0	1.34	55,228
PC1: # Trades	-2.84	22.97	-0.52	0	1.37	55,228
PC1: Both	-3.09	16.14	-0.73	0	1.76	55,228

Table 3. Fake trading and exchange characteristics. This table reports comparisons of mean principal-components-based measures of fake trading for groups of high/low exchange characteristics associated with regulation/compliance/transparency and the differences between the two groups. In columns 1-3, the fake trading measure, PC1: Volume, is the first principal component based on trading-volume-based measures. In columns 4-6, the fake trading measure, PC1: # Trades, is the first principal component based on number-of-trades-based measures. In columns 7-9, the fake trading measure, PC1: Both, is the first principal component based on both trading-volume-based and number-of-trades-based measures. In case of discrete variables, Low and High correspond to indicator variables equaling zero and one. In case of continuous variables, Low and High are defined based on whether the value of the variable is higher or lower than its in-sample median. See Table A.1 in the Appendix for variable definitions. The sample is all exchange-months with non-missing values of the characteristics being compared. * Significant at 5 percent; ** Significant at 1 percent; *** Significant at 0.1 percent.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
	PC1: Volume			**PC1: # Trades**			**PC1: Both**		
	Low	**High**	**Diff.**	**Low**	**High**	**Diff.**	**Low**	**High**	**Diff.**
AML	0.197	-0.085	-0.281 ***	0.293	-0.126	-0.419 ***	0.375	-0.162	-0.537 ***
			(0.012)			(0.013)			(0.016)
KYC	0.077	-0.135	-0.212 ***	0.029	-0.016	-0.045 ***	0.031	-0.055	-0.086 ***
			(0.012)			(0.012)			(0.016)
Crypto-friendly location	-0.546	0.180	0.726 ***	-0.552	0.182	0.734 ***	-0.789	0.260	1.048 ***
			(0.013)			(0.013)			(0.017)
Bad news	-0.333	0.236	0.568 ***	-0.435	0.308	0.744 ***	-0.559	0.396	0.955 ***
			(0.011)			(0.011)			(0.015)
Multiplatform	-0.199	0.382	0.581 ***	-0.415	0.794	1.209 ***	-0.461	0.883	1.344 ***
			(0.012)			(0.011)			(0.015)
Web popularity	0.071	-0.159	-0.230 ***	0.209	-0.249	-0.458 ***	0.215	-0.302	-0.517 ***
			(0.011)			(0.012)			(0.015)
Twitter	0.348	-0.279	-0.627 ***	0.441	-0.354	-0.796 ***	0.587	-0.471	-1.058 ***
			(0.011)			(0.011)			(0.014)
Reddit	0.246	-0.253	-0.499 ***	0.348	-0.358	-0.706 ***	0.448	-0.461	-0.909 ***
			(0.011)			(0.011)			(0.014)
Github	0.040	-0.173	-0.212 ***	0.086	-0.374	-0.460 ***	0.094	-0.411	-0.506 ***
			(0.015)			(0.015)			(0.019)

Table 4. Alternative fake trading measures. This table reports estimates of regressions of alternative, correlation-based measures of fake trading on the three principal-components-based measures of fake trading. The dependent variable in columns 1 and 2 is trading-volume correlation-based measures. To compute it, each month for each exchange and currency pair that is traded on at least five exchanges, we calculate the correlation between the volume of trading on that exchange during ten-minute intervals and contemporaneous aggregate trading volume on all exchanges. Number-of-trades correlation-based measure, used as the dependent variable in columns 3 and 4, and price correlation-based measure, used as the dependent variable in columns 5 and 6, are computed similarly. In the upper part of the table, the main independent variable is PC1: Volume, the first principal component based on trading-volume-based measures. In the middle part of the table, the main independent variable is PC1: # Trades, the first principal component based on number-of-trades-based measures. In the lower part of the table, the main independent variable is PC1: Both, the first principal component based on both trading-volume-based and number-of-trades-based measures. In even columns we control for Web popularity of the exchange, measured as one minus the ratio of the highest Alexa rank of exchanges in the sample divided by the exchange's Alexa rank, and for the exchange's social media presence, proxied by the natiral logarithm of the number of Reddit posts and Twitter tweets. See Table A.1 in the Appendix for variable definitions. The sample is all exchange-currency pair-month observations with non-missing values of dependent and independent variables. The regressions are estimated at the exchange-currency pair-month level. We report robust standard errors in parenthesis. * Significant at 5 percent; ** Significant at 1 percent; *** Significant at 0.1 percent.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent variable	**Corr (Volume)**		**Corr (# Trades)**		**Corr (Price)**	
PC1: Volume	-0.007 ***	-0.006 ***	-0.009 ***	-0.003 *	-0.001	0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Web popularity	No	Yes	No	Yes	No	Yes
Social media	No	Yes	No	Yes	No	Yes
# Obs.	26,630	16,537	26,630	16,537	26,630	16,537
Adj. R^2	0.083	0.142	0.092	0.186	0.054	0.060
PC1: # Trades	-0.018 ***	-0.007 **	-0.015 ***	-0.005 *	0.001	-0.000
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Web popularity	No	Yes	No	Yes	No	Yes
Social media	No	Yes	No	Yes	No	Yes
# Obs.	26,630	16,537	26,630	16,537	26,630	16,537
Adj. R^2	0.091	0.141	0.095	0.186	0.050	0.060
PC1: Both	-0.011 ***	-0.004 *	-0.010 ***	-0.006 **	-0.001	-0.000
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Web popularity	No	Yes	No	Yes	No	Yes
Social media	No	Yes	No	Yes	No	Yes
# Obs.	26,630	16,537	26,630	16,537	26,630	16,537
Adj. R^2	0.087	0.141	0.094	0.186	0.052	0.061

Table 5. Chinese Ban. This table reports estimates of regressions of the three principal components-based measures of fake trading around the ban in China of crypto exchanges in September 2017 on Treated exchange indicator, Post-ban indicator, and their interaction. The sample in columns 1-3 includes all exchanges that operated in China prior to the ban and moved to Pacific Asian countries as a result of the ban (treated exchanges) and exchanges that operated in Pacific Asian countries prior to the ban (control exchanges). The sample in columns 4-6 includes all exchanges that operated in China prior to the ban and moved to Hong Kong as a result of the ban (treated exchanges) and exchanges that operated in Hong Kong prior to the ban (control exchanges). The sample period is the first three quarters of 2017 and the first two quarters of 2018. In columns 1 and 4, the dependent variable is PC1: Volume, the first principal component based on trading-volume-based measures. In columns 2 and 5, the dependent variable is PC1: # Trades, the first principal component based on number-of-trades-based measures. In columns 3 and 6, the dependent variable is PC1: Both, the first principal component based on both trading-volume-based and number-of-trades-based measures. Treated is an indicator variable equaling one for treated firms. Post-ban equals one for the first two quarters of 2018. Treated × Post ban is the interaction between these two variables. We control for exchange characteristics and currency pair characteristics described in detail in Table 6. See Table A.1 in the Appendix for variable definitions. The regressions are estimated at the exchange-currency pair-month level. We report robust standard errors in parentheses. * Significant at 5 percent; ** Significant at 1 percent; *** Significant at 0.1 percent.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Sample	**Pacific Asia (Treated exchanges: 7) (Control exchanges: 11)**			**Hong Kong (Treated exchanges: 5) (Control exchanges: 5)**		
Dependent variable	**PC1: Volume**	**PC1: # Trades**	**PC1: Both**	**PC1: Volume**	**PC1: # Trades**	**PC1: Both**
Treated	1.679 ***	1.379 ***	1.596 ***	1.662 ***	0.833 ***	1.956 ***
	(0.046)	(0.043)	(0.056)	(0.115)	(0.096)	(0.140)
Post-ban	-0.378 ***	-0.420 ***	-0.594 ***	-0.356 ***	-0.306 ***	-0.473 ***
	(0.070)	(0.065)	(0.085)	(0.074)	(0.062)	(0.090)
Treated × Post-ban	-1.342 ***	-1.297 ***	-1.146 ***	-0.791 ***	-0.745 ***	-1.154 ***
	(0.074)	(0.069)	(0.090)	(0.208)	(0.173)	(0.253)
Exchange characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Currency pair characteristics	Yes	Yes	Yes	Yes	Yes	Yes
# Obs.	6,554	6,554	6,554	3,524	3,524	3,524
Adj. R^2	0.180	0.213	0.224	0.182	0.226	0.213

Table 6. Static competition and fake trading. This table reports estimates of regressions of three principal components-based measures of fake trading on exchange-level and currency-pair-level characteristics. In columns 1 and 2, the dependent variable is PC1: Volume, the first principal component based on trading-volume-based measures. In columns 3 and 4, the dependent variable is PC1: # Trades, the first principal component based on number-of-trades-based measures. In columns 5 and 6, the dependent variable is PC1: Both, the first principal component based on both trading-volume-based and number-of-trades-based measures. See Table A.1 in the Appendix for definitions of exchange and currency-pair characteristics. The set of independent variables includes base pair (BTC, ETH, USDT) fixed effects and year-quarter fixed effects. In odd columns, the regressions include the geographical region of exchange location. In even columns, exchange fixed effects are included. The sample is all exchange-currency pair-month observations with non-missing values of dependent and independent variables. The regressions are estimated at the exchange-currency pair-month level. We report robust standard errors in parentheses. * Significant at 5 percent; ** Significant at 1 percent; *** Significant at 0.1 percent.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent variable	**PC1: Volume**		**PC1: # Trades**		**PC1: Both**	
Exchange characteristics						
log (# Currency pairs)	0.044 ***		0.156 ***		0.080 ***	
	(0.011)		(0.010)		(0.013)	
Market share: Volume	-0.034 ***		-0.049 ***		-0.058 ***	
	(0.005)		(0.004)		(0.006)	
log (Age)	-0.043 ***		-0.028 ***		(0.002)	
	(0.009)		(0.008)		(0.011)	
Location						
Africa	1.060 ***		0.697 ***		1.317 ***	
	(0.054)		(0.049)		(0.066)	
Asia (other than China)	0.815 ***		0.264 ***		0.733 ***	
	(0.033)		(0.030)		(0.039)	
China	1.567 ***		2.047 ***		2.679 ***	
	(0.035)		(0.032)		(0.042)	
Central and South America	1.386 ***		1.221 ***		1.924 ***	
	(0.038)		(0.034)		(0.046)	
Eastern Europe	0.506 ***		-0.632 ***		-0.132 ***	
	(0.031)		(0.028)		(0.038)	
Western Europe	0.332 ***		-0.148 ***		0.127 ***	
	(0.029)		(0.026)		(0.035)	
Europe: Islands	0.593 ***		0.011		0.407 ***	
	(0.030)		(0.028)		(0.037)	
Currency-pair characteristics						
log (Listed on # exchanges)	0.044 ***	0.048 ***	0.025 ***	0.022 ***	0.046 ***	0.048 ***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
HHI Currency pair across exchanges: Volume	-0.116 ***	-0.211 ***	-0.125 ***	-0.204 ***	-0.117 ***	-0.218 ***
	(0.023)	(0.022)	(0.020)	(0.020)	(0.027)	(0.026)
log (Age of listing on the exchange)	0.029	0.050	-0.095 **	(0.038)	(0.070)	(0.013)
	(0.035)	(0.033)	(0.031)	(0.030)	(0.042)	(0.040)
Token	-0.091 ***	-0.060 ***	-0.126 ***	-0.101 ***	-0.153 ***	-0.112 ***
	(0.011)	(0.011)	(0.010)	(0.010)	(0.013)	(0.013)
Exchange FE	no	yes	no	yes	no	yes
Base pair FE	yes	yes	yes	yes	yes	yes
Year-quarter FE	yes	yes	yes	yes	yes	yes
# Obs.	55,220	55,228	55,220	55,228	55,220	55,228
Adj. R^2	0.146	0.221	0.334	0.389	0.282	0.347

Table 7. Dynamic competition and fake trading. This table reports estimates of regressions of three principal components-based measures of fake trading on measures of competition, indicators of entry and exit by competitor exchanges, and the interaction of these entry/exit indicators with high competition indicator. In columns 1, 4, and 7, the dependent variable is PC1: Volume, the first principal component based on trading-volume-based measures. In columns 2, 6, and 8, the dependent variable is PC1: # Trades, the first principal component based on number-of-trades-based measures. In columns 3, 6, and 9, the dependent variable is PC1: Both, the first principal component based on both trading-volume-based and number-of-trades-based measures. In columns 1-3, "general" competitors are a set of all exchanges in our dataset. In columns 4-6, "geographical" competitors are a subset of exchanges that operate in the same geographical region as the focal exchange, where regions are defined as in Panel C of Table 1. In columns 7-9, "Operational" competitor is the exchange belonging to the set of general competitors that has the largest overlap of pairs listed with the focal exchange. Moderate competition is an indicator equaling one if a currency pair is listed on at least two and at most seven exchanges. High competition is an indicator equaling one if a currency pair is listed on at least eight exchanges. Competitor entry is an indicator equaling one if there exists at least one exchange that has not previously listed a currency pair starts listing it in a given month. Competitor exit is an indicator equaling one if there is at least one exchange that has listed a currency pair last month and does not list it in a given month. We control for exchange characteristics and currency pair characteristics described in detail in Table 6. See Table A.1 in the Appendix for definitions of exchange and currency pair characteristics. The set of independent variables also includes base pair (BTC, ETH, USDT) fixed effects and year-quarter fixed effects. The sample is all exchange-currency pair-month observations with non-missing values of dependent and independent variables. The regressions are estimated at the exchange-currency pair-month level. We report robust standard errors in parentheses. * Significant at 5 percent; ** Significant at 1 percent; *** Significant at 0.1 percent.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Competitor type		**General**			**Geographical**			**Operational**	
Dependent Variable	**PC1: Volume**	**PC1: # Trades**	**PC1: Both**	**PC1: Volume**	**PC1: # Trades**	**PC1: Both**	**PC1: Volume**	**PC1: # Trades**	**PC1: Both**
Moderate competition	0.062 *	0.184 ***	0.197 ***	0.060 *	0.178 ***	0.192 ***	0.045	0.162 ***	0.170 ***
	(0.028)	(0.028)	(0.036)	(0.028)	(0.028)	(0.036)	(0.028)	(0.028)	(0.036)
High competition	0.437 ***	0.417 ***	0.622 ***	0.420 ***	0.386 ***	0.588 ***	0.438 ***	0.413 ***	0.618 ***
	(0.043)	(0.043)	(0.055)	(0.043)	(0.043)	(0.055)	(0.042)	(0.042)	(0.054)
Competitor entry	0.137 ***	0.195 ***	0.223 ***	0.167 ***	0.203 ***	0.269 ***	0.196 ***	0.197 ***	0.291 ***
	(0.023)	(0.023)	(0.029)	(0.026)	(0.026)	(0.033)	(0.034)	(0.034)	(0.043)
Competitor exit	-0.177 ***	-0.291 ***	-0.342 ***	0.017	0.078	0.043	0.062	0.064	0.021
	(0.037)	(0.037)	(0.048)	(0.047)	(0.047)	(0.061)	(0.065)	(0.065)	(0.083)
Competitor entry	0.084	0.031	0.071	0.131 **	0.153 ***	0.193 ***	0.312 ***	0.138 *	0.300 **
× High competition	(0.043)	(0.043)	(0.055)	(0.045)	(0.045)	(0.058)	(0.048)	(0.075)	(0.096)
Competitor exit	-0.203 ***	-0.193 **	-0.283 ***	0.117	0.023	0.081	-0.119	-0.472 ***	-0.489 **
× High competition	(0.059)	(0.059)	(0.076)	(0.068)	(0.069)	(0.088)	(0.138)	(0.139)	(0.177)
Exchange characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Currency-pair characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Base pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Obs.	55,220	55,220	55,220	55,220	55,220	55,220	55,220	55,220	55,220
Adj. R^2	0.083	0.115	0.117	0.084	0.115	0.117	0.083	0.114	0.116

Table 8. Effects of fake trading on trading volume. This table reports estimates of regressions of trading volume on lagged trading volume, as well as contemporaneous and lagged principal components-based measures of fake trading. In columns 1-2, 5-6, and 7-8 the dependent variable is trading volume and the regressions are estimated using OLS. Columns 4, 8, and 12 present estimates of the second stage of 2SLS regressions, in which the dependent variable is trading volume and lagged volume is replaced a fitted value from the first-stage regressions, whose estimates are reported in columns 3, 7, and 11 respectively. The dependent variable in the first-stage regressions is lagged volume and the independent variables are lagged fake trading measure, lagged average Bitcoin price throughout the month of the observation and lagged squared Bitcoin price. In columns 1-4, the fake trading measure is PC1: Volume, the first principal component based on trading-volume-based measures. In columns 5-8, the fake trading measure is PC1: # Trades, the first principal component based on trading-volume-based measures. In columns 9-12, the fake trading measure is PC1: Both, the first principal component based on both trading-volume-based and number-of-trades-based measures. We control for exchange characteristics and currency pair characteristics described in detail in Table 6. See Table A.1 in the Appendix for definitions of exchange and currency pair characteristics. The set of independent variables also includes base pair (BTC, ETH, USDT) fixed effects and year-quarter fixed effects. The sample is all exchange-currency pair-month observations with non-missing values of dependent and independent variables. The regressions are estimated at the exchange-currency pair-month level. We report robust standard errors in parentheses. * Significant at 5 percent; ** Significant at 1 percent; *** Significant at 0.1 percent.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
Estimation	**OLS**	**OLS**	**2SLS 1st stage**	**2SLS 2nd stage**	**OLS**	**OLS**	**2SLS 1st stage**	**2SLS 2nd stage**	**OLS**	**OLS**	**2SLS 1st stage**	**2SLS 2nd stage**
Dependent Variable	**Volume**	**Volume**	**Lag (Volume)**	**Volume**	**Volume**	**Volume**	**Lag (Volume)**	**Volume**	**Volume**	**Volume**	**Lag (Volume)**	**Volume**
PC1: Volume	0.155 ***	0.081 ***		0.152 ***								
	(0.005)	(0.002)		(0.005)								
PC1: # Trades					0.225 ***	0.132 ***		0.224 ***				
					(0.006)	(0.003)		(0.006)				
PC1: Both									0.193 ***	0.116 ***		0.192 ***
									(0.005)	(0.002)		(0.005)
Lag (PC1: Volume)	0.106 ***	-0.054 ***	0.311 ***	-0.036								
	(0.005)	(0.002)	(0.003)	(0.019)								
Lag (PC1: # Trades)					0.155 ***	-0.084 ***	0.404 ***	-0.028				
					(0.005)	(0.003)	(0.003)	(0.022)				
Lag (PC1: Both)									0.122 ***	-0.080 ***	0.354 ***	-0.063 **
									(0.005)	(0.002)	(0.003)	(0.020)
Lag (Volume)		0.902 ***		0.454 ***		0.891 ***		0.455 ***		0.893 ***		0.524 ***
		(0.002)		(0.058)		(0.002)		(0.053)		(0.002)		(0.056)
Lag (BTC Price)			0.028 ***				0.028 ***				0.025 ***	
			(0.001)				(0.001)				(0.001)	
Lag (BTC Price^2)			-0.001 ***				-0.001 ***				-0.001 ***	
			0.000				0.000				0.000	
Exchange characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Currency-pair characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Base pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Obs.	49,861	49,861	49,861	49,861	49,861	49,861	49,861	49,861	49,861	49,861	49,861	49,861
Adj. R^2	0.365	0.869	0.178	0.366	0.449	0.873	0.291	0.449	0.432	0.873	0.272	0.433

Table 9. Effects of fake trading on web popularity and estimated exchange revenue. This table reports estimates of regressions of a measure of web popularity and of estimated revenue of the exchange on trading volume over the past three months (in Panel A) and over the past twelve months (in Panel B). Web popularity is measured as one minus the ratio of the highest Alexa rank of exchanges in the sample divided by the exchange's Alexa rank. See Section 6.2 for details of estimation of exchange's revenue. In columns 1 and 4, the fake trading measure is PC1: Volume, the first principal component based on trading-volume-based measures. In columns 2 and 5, the fake trading measure is PC1: # Trades, the first principal component based on trading-volume-based measures. In columns 3 and 6, the fake trading measure is PC1: Both, the first principal component based on both trading-volume-based and number-of-trades-based measures. We control for exchange characteristics, described in detail in Table 6 and defined in Table A.1 in the Appendix. The set of independent variables also includes year-quarter fixed effects. The sample is all exchange-month observations with non-missing values of dependent and independent variables. The regressions are estimated at the exchange-month level. We report robust standard errors in parentheses. * Significant at 5 percent; ** Significant at 1 percent; *** Significant at 0.1 percent.

Panel A: Short-term effects

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent variable		**Web popularity**			**Estimated exchange revenue**	
Lag3 (PC1: Volume)	1.778 **			0.496 ***		
	(0.624)			(0.028)		
Lag3 (PC1: # Trades)		1.892 **			0.511 ***	
		(0.632)			(0.028)	
Lag3 (PC1: Both)			1.481 **			0.414 ***
			(0.483)			(0.020)
Exchange characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
# Obs.	722	722	722	724	724	724
Adj. R^2	0.467	0.48	0.511	0.475	0.498	0.497

Panel B: Long-term effects

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent variable		**Web popularity**			**Estimated exchange revenue**	
Lag12 (PC1: Volume)	-3.115 ***			-0.210 ***		
	(0.820)			(0.037)		
Lag12 (PC1: # Trades)		-3.425 ***			-0.259 ***	
		(0.808)			(0.036)	
Lag12 (PC1: Both)			-2.700 ***			-0.201 ***
			(0.630)			(0.028)
Exchange characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
# Obs.	428	428	428	430	430	430
Adj. R^2	0.649	0.646	0.646	0.717	0.719	0.719

Appendix: Variable definitions

Table A.1. Variable description

Variable	Type	Description
Market		
Market cap: CMC	Continuous $(0, \infty)$	Aggregate monthly market capitalization at the end of a month in millions of $U.S. Dollars as per www.CoinMarketCap.com
MarketCap: Kaiko	Continuous $(0, \infty)$	Aggregate monthly market capitalization at the end of a month in millions of $U.S. Dollars as per www.Kaiko.com
Market cap CMC / Market Cap Kaiko	Continuous $(0, 1]$	The ratio of Market cap CMC and Market cap Kaiko
Currencies: All	Integer	Total number of crypto currencies (tokens and coins) listed on at least one exchange in a given month
Currencies: Tokens	Integer	Total number of crypto tokens listed on at least one exchange in a given month
Currencies: Coins	Integer	Total number of crypto coins listed on at least one exchange in a given month
Currencies: Entry	Integer	Total number of crypto currencies listed on at least one exchange in a given month and not listed on any exchange in the previous month
Currencies: Exit	Integer	Total number of crypto currencies not listed on any exchange in a given month and listed on at least one exchange in the previous month
Currency pairs: All	Integer	Total number of distinct currency pairs listed on at least one exchange in a given month
Currency pairs: Tokens	Integer	Total number of distinct currency pairs involving a token as a quote currency listed on at least one exchange in a given month
Currency pairs: Coins	Integer	Total number of distinct currency pairs involving a coin as a quote currency listed on at least one exchange in a given month
Currency pairs: Entry	Integer	Total number of currency pairs listed on at least one exchange in a given month and not listed on any exchange in the previous month
Currency pairs: Exit	Integer	Total number of currency pairs not listed on any exchange in a given month and listed on at least one exchange in the previous month
Volume: $U.S. (MM)	Continuous $(0, \infty)$	Aggregate reported volume of trading in all currency pairs on all exchanges in a given month in billions of $U.S.
Volume: BTC (M)	Continuous $(0, \infty)$	Aggregate reported volume of trading in all currency pairs on all exchanges in a given month in millions of BTC
# Trades (M)	Integer	Aggregate reported number of trades in all currency pairs on all exchanges in a given month in millions
Exchanges	Integer	Number of exchanges listing at least one currency pair in a given month
Exchanges: Entry	Integer	Number of exchanges listing at least one currency pair in a given month and not listing any currency pairs in the previous month
Exchanges: Exit	Integer	Number of exchanges not listing any currency pairs in a given month and listing at least one currency pair in the previous month
HHI Exchanges: # Currencies	Continuous $(0, 1]$	Sum of squared number of currencies listed on each exchange divided by the squared sum of number of currencies listed on all exchanges in a given month
HHI Exchanges: Volume	Continuous $(0, 1]$	Sum of squared reported trading volume on each exchange divided by the squared sum of reported trading volume on all exchanges in a given month
HHI Exchanges: # Trades	Continuous $(0, 1]$	Sum of squared reported number of trades on each exchange divided by the squared sum of reported number of trades on all exchanges in a given month

Table A.1. Variable description – continued

Variable	Type	Description
Currency pairs		
Listed on # exchanges:	Integer	The number of exchanges on which a currency pair is listed in a given month
log (Listed on # exchanges):	Continuous $(0, \infty)$	Natural logarithm of the number of exchanges on which a currency pair is listed in a given month
HHI: Currency pair across exchanges: Volume	Continuous $(0, 1]$	Squared reported volume of trading in a currency pair on each exchange divided by the squared sum of total reported volume of trading in the currency pair on all exchange in a given month
HHI: Currency pair across exchanges: # Trades	Continuous $(0, 1]$	Squared reported number of trades involving a currency pair on each exchange divided by the squared sum of total reported number of trades involving the currency pair on all exchange in a given month
Age of listing on any exchange	Integer	Difference in months between current month and the first month a currency pair was listed on any exchange
Age of listing on a given exchange	Integer	Difference in months between current month and the first month a currency pair was listed on a particular exchange
Time to listing	Integer	Difference in months between the first month a currency pair was listed on a particular exchange and the first month the currency pair was listed on any exchange
Token	Indicator	Equals one for tokens issued in an ICO

Table A.1. Variable description – continued

Variable	Type	Description
Exchanges		
Age	Integer	The difference in month between current month and the first month any currency pair was listed on an exchange
Market share: Volume	Continuous $(0, 1]$	Ratio of reported aggregate trading volume on a given exchange and total reported trading volume on all exchanges in a given month
Market share: # Trades	Continuous $(0, 1]$	Ratio of reported aggregate number of trades on a given exchange and total reported number of trades on all exchanges in a given month
Market share: Currency pairs	Continuous $(0, 1]$	Ratio of the number of currency pairs listed on a given exchange and the sum across all exchanges of the numbers of currency pairs listed on them in a given month
Currency pairs	Integer	Number of currency pairs listed on an exchange in a given month
Currency pairs: Entry	Integer	Number of currency pairs that are listed on an exchange in a given month that were not listed on the exchange in the previous month
Currency pairs: Exit	Integer	Number of currency pairs that are not listed on an exchange in a given month that were listed on the exchange in the previous month
AML	Indicator	Equals one if an exchange implemented an AML policy and provides detailed information about conformity with accepted international AML procedures
KYC	Indicator	Equals one if there are evidence that the exchange provides clear guidelines, requires documents, and verifies sources of clients' funds
Crypto-friendly location	Indicator	Equals one if an exchange is located in Singapore, Russia, Estonia, Malta, Luxembourg or Switzerland
Bad news	Indicator	Equals one if an exchange has negative news in the current quarter. News are considered bad if it is related to hack attacks, poor review results, scams or theft.
Multiplatform	Indicator	Equals one if an exchange has both a regular platform and a decentralized platform
Alexa (K)	Integer	Rank in thousands of an exchange's website, as reported by www.Alexa.com
Reddit	Integer	Number of Reddit posts from an exchange's official Reddit account in a given month
Twitter	Integer	Number of Twitter tweets from an exchange's official Twitter account in a given month
Github	Integer	The amount of code reviews (commits) at the exchange main Github repository in a given month
Exchange-currency pairs		
Volume: $U.S. (M)	Continuous $(0, \infty)$	Aggregate reported volume of trading in all currency pairs on a given exchange in a given month in millions of $U.S.
Volume: BTC (K)	Continuous $(0, \infty)$	Aggregate reported volume of trading in all currency pairs on a given exchange in a given month in thousands of BTC
# Trades	Integer	Aggregate reported number of trades involving all currency pairs on a given exchange in a given month in thousands
HHI Currency pairs within exchange: Volume	Continuous $(0, 1]$	Sum of squared reported trading volume of all currency pairs listed on a given exchange divided by the squared sum of aggregate reported trading volume of all currency pairs in the exchange in a given month
HHI Currency pairs within exchange: # Trades	Continuous $(0, 1]$	Sum of squared reported number of trades involving all currency pairs listed on a given exchange divided by the squared sum of aggregate reported number of trades involving all currency pairs in the exchange in a given month

Table A.1. Variable description – continued

Variable	Type	Description
Quality measures		
MAD: Volume	Continuous (0, 1)	Mean absolute distance between the frequency of leading digits of the trading volume series in a currency pair on a given exchange in a given month and the frequency of leading digits given by Benford's Law
MAD: # Trades	Continuous (0, 1)	Mean absolute distance between the frequency of leading digits of the number of trades series in a currency pair on a given exchange in a given month and the frequency of leading digits given by Benford's Law
KS: Volume	Continuous [0, 1]	Kolmogorov-Smirnov distance between the cumulative distribution function (c.d.f.) of the natural logarithm of trading volume in a currency pair on a given exchange in a given month and the c.d.f. of standard normal distribution
KS: # Trades	Continuous [0, 1]	Kolmogorov-Smirnov distance between the cumulative distribution function (c.d.f.) of the natural logarithm of the number of trades involving a currency pair on a given exchange in a given month and the c.d.f. of standard normal distribution
EDM: Volume	Integer	The number of structural breaks in series of trading volume within ten-minutes intervals of a currency pair on a given exchange in a given month identified by E-Divisive with Medians algorithm
EDM: # Trades	Integer	The number of structural breaks in series of number of trades within ten-minutes intervals involving a currency pair on a given exchange in a given month identified by E-Divisive with Medians algorithm
Principal components		
PC1: Volume	Continuous	The first principle component of Mad: Volume, KS: Volume, and EDM: Volume
PC1: # Trades	Continuous	The first principle component of Mad: # Trades, KS: # Trades, and EDM: # Trades
PC1: Both	Continuous	The first principle component of Mad: Volume, KS: Volume, EDM: Volume, Mad: # Trades, KS: # Trades, and EDM: # Trades