What Drives the Covariation of Cryptocurrency Returns?

AMIN SHAMS*

December 27, 2019

Abstract

Demand for cryptocurrencies not only signals investment motives but also can be a sign of user adoption that affects the fundamental value of these assets. Therefore, common demand shocks can be the primary driver of cryptocurrency return structure. This paper constructs a "connectivity" measure proxying for common demand shocks and shows that it explains substantial covariation in the cross-section of cryptocurrency returns. Cryptocurrencies trade on more than 150 exchanges, and due to geographical and other restrictions, these exchanges serve different investor clienteles who show correlated demand across currencies listed on the same exchange. I define connectivity as the degree of overlap in the set of exchanges two cryptocurrencies trade on. I show that connected currencies exhibit substantial contemporaneous covariation. In addition, currencies connected to those that perform well outperform currencies connected to those that perform poorly by 71 basis points over the next day and 214 basis points over the next week. Evidence from new exchange listings and a quasi-natural experiment exploiting the shutdown of Chinese exchanges shows that the results cannot be explained by endogenous sorting of currencies into exchanges. Using machine learning techniques to analyze social media data, I find that the demand effects are 40 to 50% larger for currencies that rely more heavily on network externalities of user adoption. This amplified effect is consistent with the notion that demand for a cryptocurrency originates not only from investment motives but also the user adoption that can affect the fundamental value of these assets.

*I am very grateful for the advice and guidance I have received from members of my dissertation committee: John M. Griffin (co-chair), Sheridan Titman (co-chair), Carlos Carvalho, Cesare Fracassi, Travis Johnson, and Samuel Kruger. Helpful comments were received from Hassan Afrouzi, Aydogan Alti, Tin Dinh, Iman Dolatabadi, Michael Sockin, Aleh Tsyvinski, and seminar participants at George Mason University, Harvard Business School, Rice University, the Ohio State University, the University of California, Los Angeles, the University of Florida, the University of Iowa, the University of North Carolina at Chapel Hill, the University of Toronto, the University of Virginia, and Tulane University. Data for the project were provided by Integra FEC. I would also like to acknowledge financial support in the form of a doctoral fellowship from the Blockchain Initiative at McCombs School of Business, partly funded by a generous donation from Ripple Labs Inc.
Fisher College of Business, Ohio State University.
Email: shams.22@osu.edu

The number of cryptocurrencies has soared rapidly in recent years, and there are now nearly 5,000 tradable cryptocurrencies in the market. The majority of these digital currencies aim to either compete with fiat money or provide an alternative funding source and native medium of exchange for platforms that offer specific products. Beyond the hype and controversy surrounding the market, there is an actual trading ecosystem for cryptocurrencies. There exist over 150 crypto-exchanges around the world with sizable trading activity and real money involved. Due to their unique features, extensive speculation in the market, and potential implications for capital allocation, academics have recently become interested in the asset pricing properties of cryptocurrencies. Yet, our understanding of the structure and drivers of returns in this market remains quite limited. This paper takes a step towards better understanding the drivers of returns in this market and establishing a foundation for future research in this area.

I start by documenting a persistent comovement structure in cryptocurrencies. Panel A of Figure 1 shows the distribution of pairwise correlation of daily returns for the largest 50 cryptocurrencies. The figure shows a wide variation ranging from zero correlation for some pairs to nearly 0.8 for the others. Importantly, Panel B of Figure 1 shows that this correlation structure is persistent. Pairs that correlate highly this month exhibit a high correlation next month as well, and those with a low correlation for this month also show a low correlation next month. The central question of this paper is what forces explain this persistent return structure. Is the high correlation mainly driven by similarity in products or characteristics such as the underlying technology and size, or are there other primary drivers of prices in this market?

I find that while similarity in characteristics such as technology and size explain some covariation in cryptocurrencies, commonalities in demand shocks is the main driver of this structure. I proxy for demand commonalities using the trading exchange of cryptocurrencies. Cryptocurrencies trade on numerous exchanges, more than 150 of which are included in my data. Due to geographical and other restrictions, these exchanges attract different investor clientele. For example, a South Korean exchange named Bithumb is only open to South Korean investors. Zaif is an exchange popular among Japanese traders because it only accepts deposits in Japanese Yen. Importantly, different cryptocurrencies demonstrate different levels of trading on different exchanges. For example, Ripple, the third largest cryptocurrency, trades heavily on South Korean exchanges during my sample period. However, a much smaller share of other large currencies such as Bitcoin and Ethereum traded on these exchanges. I exploit this structure to proxy for exposure of different cryptocurrencies to different investor clienteles and create a pairwise "connectivity" measure that captures similarity in their investor bases. I find that cryptocurrencies with similar investor bases show substantial comovement in order flows and returns, much larger than what can be explained by all other characteristics and technological features.

While the effect of demand pressures on prices exists in traditional asset markets, the unique features of cryptocurrencies can largely amplify this effect. Demand shocks can affect the "fundamental" value of cryptocurrencies, or the perception of investors from fundamentals. The underlying value of cryptocurrencies depends on network externalities that are driven by the participation of users and developers on the cryptocurrency platforms [Sockin and Xiong (2018) and Cong, Li, and Wang (2018)]. Buying or selling pressures can affect the fundamental value of cryptocurrencies if the pressure arises from potential users. Moreover, because utilizing the features of a cryptocurrency ecosystem often necessitates holding the token,[1] the user base is inherently interwoven with the investor base, and even pure speculative demand can be perceived as a sign of user adoption.[2] This way of thinking about the underlying value is also common in the crypto community. Discussions in media and online forums reveals that not only does the crypto community and traders consider user adoption and community building as a main source of the cryptocurrencies' value, but they also perceive buying pressures on crypto exchanges as a sign of user adoption. Here, I analyze data from currency-specific pages on social media and capture variation in the reliance of different crypto communities on the network effect. I find that demand effects are significantly stronger for cryptocurrencies that rely heavily on the network externalities.

The empirical setting exploits rich exchange-level trading data on a wide cross-section of cryp-

[1]For example, to develop decentralized applications on the EOS platform, one of the largest cryptocurrency ecosystems, participants need to hold EOS tokens.

[2]Ample anecdotal evidences from media attributed the rise in demand for Bitcoin during the boom of 2017 to global user adoption at the time.

tocurrencies merged with technical characteristics and relevant social media content. To proxy for the degree of overlapping exposure to demand shocks, I create a pairwise "connectivity" measure based on cryptocurrencies' trading locations. Cryptocurrencies are often cross-listed on multiple exchanges. Given the correlated demand on these exchanges, two currencies that trade heavily on the same set of exchanges ("connected" currencies) are exposed to similar demand shocks. The measure takes a value between zero and one, where one indicates that the share of monthly volume of two currencies is perfectly identical across different trading platforms, zero indicates they trade in entirely different locations, and a value in between indicates the extent of partial overlap.

Using the connectivity measure as a proxy for overlapping demand is motivated by three features of the cryptocurrencies' trading environment. First, due to geographical restrictions and different governance rules, there is heterogeneity in the types of investors sorted into different exchanges. Second, these exchanges are segregated due to barriers to capital transfer and frictions to register on a large number of exchanges.[3] Therefore, demand shocks to an exchange's investor base are likely to be absorbed by currencies listed on that exchange. Third, there is wide variation in the market share of different cryptocurrencies on different exchanges, which creates cross-sectional heterogeneity in the exposure of cryptocurrencies to different investor bases and to other currencies.[4]

My main set of results examines if the connectivity measure can explain variation in cryptocurrencies' comovements. I estimate a regression of pairwise correlation of market adjusted daily returns on lagged connectivity and directly control for similarity in trading volume, size, number of exchanges listing the currency, and other characteristics. I find that two cryptocurrencies that have a one-standard-deviation higher connectivity, exhibit 0.19 standard deviations higher return correlation. The results are not driven by small and less liquid currencies and hold strongly for larger currencies. Moreover, the results are not driven by exchange-level noises or the price dif-

[3]For example, Makarov and Schoar (2018) and Kroeger and Sarkar (2017) document a sizable cross-exchange arbitrage in cryptocurrency prices and attribute this arbitrage to different investor bases and frictions in moving capital across exchanges.

[4]For example, in the first half of 2018, around 23% of the trade volume of Ripple, the third largest cryptocurrency at the time of writing, occurred on a South Korean exchange. At the same time, the exchange accounted for only 2% and 3% of trading volume of Bitcoin and Ether, the first and second largest cryptocurrencies, respectively.

ference across exchanges. A within-exchange analysis reveals that among currencies listed on the same exchange, those that are more connected (due to being connected on other exchanges) show a significantly higher comovement on that exchange. Lagged connectivity alone explains around 18% of the decile spread in the next month's realized pairwise correlations.

To formalize the pricing implications of the connectivity, I examine how the price of a cryptocurrency moves in relation to the weighted average returns of its portfolio of connected currencies. Using Fama-MacBeth regressions, I find that cryptocurrencies whose connected portfolio has one percent higher daily return exhibit 32 basis points higher contemporaneous returns in the cross-section. The effect increases monotonically with the time horizon. For instance, at a weekly horizon, the magnitude is 57 basis points.

The observation that the effect increases in time horizon is consistent with investors' underreaction to the observed comovement structure. I find that the returns of connected currencies have strong positive cross-predictability after controlling for currencies' own lagged returns. A long-short zero-cost trading strategy that buys currencies in the top decile of the previous day returns of the connected portfolio, sells those in the lowest decile, and holds the portfolio over the next day, generates a before-fee daily return of 71bp. The predictability lasts for longer horizons, is robust to different portfolio constructions, and cannot be explained by non-synchronous trading.

The observed return comovements could alternatively be attributed to specific characteristics of different cryptocurrencies. A growing literature ties the intrinsic value of cryptocurrencies to characteristics of cryptocurrencies including the underlying technology and production (mining) costs.[5] Variation in characteristics have distinct testable implications for the comovement of the returns. For example, highly computationally intensive currencies should exhibit a strong comovement due to similar exposure to news about technical advances in computing power.[6] Moreover,

[5]For example, Pagnotta and Buraschi (2018) relate the price of Bitcoin to the supply of computing resources. Hayes (2017) ties the value of Bitcoin to the marginal cost of production, and Garcia et al. (2014) calculate a lower bound on the fundamental value of Bitcoin by estimating the electricity cost of mining. In contrast, Liu and Tsyvinski (2018) show that neither proxies for supply factors nor exposures to stocks, currencies, and commodities can explain the returns of major cryptocurrencies.

[6]For example, cryptocurrencies have different "consensus mechanisms." The proof of work (POW) consensus mechanism requires continually increasing computing power and electricity, but the proof of stake (POS) algorithm requires significantly less resources. This makes POW currencies more susceptible to news about computing advancements

legal shocks or changes in people's perception of different technologies can create comovement in currencies with similar cryptographic algorithm, transaction speed, privacy level, legal requirements, and underlying token industry. Importantly, if these characteristics also determine the sorting of currencies into different exchanges, then a spurious correlation between connectivity and comovement can arise in the data.

I examine the importance of these characteristics in a similar setting. The results show that pairs of currencies with similar size comove more. Furthermore, cryptocurrencies can be broadly divided into two categories: coins and tokens.[7] I find that coins comove more with other coins, and tokens with other tokens. Additionally, coins with the same consensus mechanism and tokens in the same industry have statistically significantly higher comovement. However, the combination of trading volume, hashing algorithm, consensus mechanism, token industry, and a host of other characteristics explains only 8.5% of the decile spread in pairwise correlations, which is less than half of what can be explained by the connectivity measure.

The endogeneity of the connectivity measure is a potential concern. If cryptocurrencies with similar unobservable characteristics are more likely to be listed on the same exchanges, then it could be the underlying fundamentals, and not the demand shocks, driving the relationship between connectivity and comovement. I address this concern using two tests. First, in a stacked-cohort difference-in-differences analysis, I observe that the comovements of currencies cross-listed on a new exchange significantly increase with the incumbent currencies relative to a matched sample, whereas no pre-trend is observed. Second, I use a quasi-natural experiment that exploits an exogenous shock to the trading locations due to the shutdown of Chinese exchanges by the Chinese government. Using the pre-event implied changes in the connectivity as an instrument for the actual post-event changes, I find that as the investor bases of two currencies converge by one standard deviation, their return correlation goes up by 0.17 standard deviations. These findings suggest that exogenous variations in trading location cause statistically and economically large changes in

or electricity prices.

[7] Coins are closer to traditional currencies, whereas tokens can provide users with future access to a specific product or service.

cryptocurrencies return structure.

The question remains as to what extent the observed price structure can be attributed to the idea that demand for cryptocurrencies can be interpreted as a sign of user adoption, thereby affecting the value. Cryptocurrencies vary in the degree to which they rely on the user and developer community for growth.[8] If the price impact from demand shocks is amplified through the network effect of user adoption, the impact should be larger for currencies that derive more value from this channel. I test this implication by capturing variation in the belief of the related community surrounding each cryptocurrency.

I use comments on Reddit, a social news website, to divide the cross-section of cryptocurrencies into high- and low-network-based. Cryptocurrencies have their own subpages on Reddit, and I was able to collect 12 million currency-specific comments from these pages in a machine-readable format. I classify each of these comments based on whether they relate the underlying value of the currency to factors such as the network effect, user adoption, community building, etc. I first read and label a training sample of 10,000 comments as a training set. I then use the machine learning technique, random forest, to label the remaining 12 million comments based on the training set. To verify the robustness of the results, I also create an easily replicable measure using the frequency of the most important feature extracted from the random forest model, the word "community." For each cryptocurrency, I quantify the percentage of comments that discuss the network effect. I found that this measure demonstrates a high persistence over time and is an accurate reflection of the common belief about different cryptocurrencies. Moreover, it conforms well to the theoretical characterization of the network effect discussed in the literature. Consistent with the endogenous user adoption effect in Cong, Li, and Wang (2018), network-based currencies exhibit significantly higher volatility. I find that demand commonalities translate into 40% to 50% higher return comovement for the network-based currencies. This finding suggests that a sizable portion of the observed cryptocurrency return structure can be attributed to the perceived externalities of demand

[8] For example, Ethereum provides a platform for its community to develop and use smart contracts. The users, developers, and investors are largely interwoven on the Ethereum platform and form a "community." On the other hand, these stakeholders are not as interlinked for Ripple. Ripple primarily expands through contracted software engineers and grows revenue by signing contracts directly with large banks as their customers.

for holding the tokens.

This paper contributes to a fast-growing body of literature on cryptocurrencies. Unique features of cryptocurrencies suggest that understanding investor/user behavior can be a starting point to understanding the valuation and price movements of these assets. To the best of my knowledge, this paper is the first to examine the driving forces behind cryptocurrency comovements and to document a strong amplifying effect of the externalities of users'/investors' decisions in the market. My findings suggest that a connectivity measure summarizing the structure of investor demand has substantially larger explanatory power for the comovement of cryptocurrency returns than all other characteristics and technological features combined. My paper also contributes to an extensive literature on return comovements by showing a new channel in which the demand-driven comovement can be significantly amplified in an environment where investment decisions have externalities. The rest of the paper is organized as follows. Section I discusses the institutional background and the literature. Section II overviews the testable hypotheses, the empirical design, and data. Section III describes the correlated demand within different exchanges. Section IV examines the comovement structure of the returns. Section V addresses the endogeneity concern. Section VI analyzes the network externality effect. Section VII concludes.

I. Institutional Background and Related Literature

I.A Institutional Background

To motivate the empirical design, this section describes the institutional details of the cryptocurrencies' trading environment based on which I construct the measure of similarity in the investor base. First, I highlight variation in governance rules and restrictions that are likely to attract different types of investors to different trading platforms. Second, I discuss capital barriers and other frictions that lead to segregation of these platforms. Finally, I examine variation in the market share of different cryptocurrencies on different platforms that is the source of cross-sectional heterogeneity in my setting.

I.A.1 Variation in Cryptocurrencies Trading Platforms

Cryptocurrencies trade on more than a hundred non-integrated trading platforms. My sample contains data on 74 of the most liquid ones. Despite their similarity at first glance, these exchanges differ significantly in many dimensions that may appeal to different types of investors. The main differences are discussed below.

Geographical Restrictions. Figure 2 shows that there is a wide dispersion in the geographical distribution of cryptocurrency exchanges. These exchanges are registered in nearly 30 different countries, with a large number of them in the US, China, UK, Japan, South Korea, and Russia. Variation in the country of registration is commonly accompanied by restrictions on market participants. For example, Bithumb, a South Korean exchange, only accepts South Korean citizens. The major Hong Kong-based exchange, Bitfinex, terminated relationships with U.S. banks, therefore making it much more difficult for U.S. customers to trade on the exchange. OKEX, another exchange registered in Hong Kong, does not accept customers from certain countries including the United States and Hong Kong itself. Regulators in Japan and Hong Kong warned investors against using Binance, an exchange initially registered in China. Restrictions like these apply to many other exchanges, which can lead to a wide heterogeneity in the home country of investors on different trading platforms.

Identity Verification Requirements. Different exchanges require different levels of identity verifications. Some exchanges, usually those registered in locations such as the Republic of Seychelles and Cayman Islands, have minimal legal restrictions. Other exchanges have strict requirements. For example, Coinbase, a US-based exchange, agreed to share customers' information with the Internal Revenue Service (IRS) for tax purposes. Gemini, another exchange based in the US, requires the Social Security Number (SSN) of customers. Different verification requirements can attract different types of investors. For example, an institutional investor may choose to trade on an exchange with stricter restrictions to avoid counterparty risk, while a large cryptocurrency whale who prefers to keep his or her identity secret is likely to use platforms with fewer requirements.

Limitations on Deposits, Withdrawal, and Use of Fiat Currencies. Some platforms allow trading with fiat currencies, while some only accept digital currencies. Platforms built on fiat-like digital currencies, such as Tether, might be more appealing to investors facing cross-country capital restrictions, while as Griffin and Shams (2018) show they might also bear the risk of redeeming these currencies into cash. There is also variation in the fiat monies supported by different exchanges that can further widen the heterogeneity in investor type on different platforms. For example, Kraken accepts multiple fiat currencies including CAD, EUR, GBP, JPY, and USD, while the exchange Zaif only accepts JPY.

Transaction Fees. Exchanges have different deposit, withdrawal, and transaction fees. The fee structure can organically sort different types of traders into different platforms. For example, low trading fee exchanges might be more tailored towards large active traders.

I.A.2 Frictions Across Trading Platforms

Frictions for moving capital across crypto exchanges can lead to segregation of these exchanges. These frictions can arise due to cross-country capital restrictions, fees, and risks associated with withdrawals and deposits from and to the exchanges as well as slow confirmation of these transactions. Moreover, costs and risks associated with trading on multiple platforms intensify this segregation. Several platforms are subject to know-your-customer (KYC) regulations, and even those that are not may require revealing sensitive information such as wallet addresses used for deposit or withdrawal. Among other costs associated with trading on multiple exchanges, the more platforms that have a customer's sensitive information, the more likely such information could be compromised. Therefore, traders might prefer to trade on a limited set of exchanges.

Such frictions create limits to arbitrage that can lead to sizable cross-exchange price differences. For example, Makarov and Schoar (2018) and Kroeger and Sarkar (2017) document a sizable price difference between cryptocurrency exchanges. Due to these frictions, shocks to the investor base in a trading platform are more likely to be absorbed by currencies listed on that platform, which leads to correlated demand shocks across currencies listed on the same exchange.

I.A.3 Variation in Share of Cryptocurrencies on Different Exchanges

There is a wide dispersion in the market share (by volume) of cryptocurrencies on different trading platforms. Large currencies such as Bitcoin and Ether trade on almost all major platforms, though to varying degrees. Other top coins, such as EOS, Bitcoin Cash, and Ripple, trade on several exchanges but not all. The market share of these currencies on different exchanges varies from 8% to 100% for Bitcoin, 0% to 49% for Ether, 0% to 36% for EOS, 0% to 22% for Bitcoin Cash, and 0% to 40% for Ripple.

Figure 3 shows that there is large heterogeneity in the volume share of major currencies (as a percentage of a currency's total volume) on major exchanges. For example, in the first half of 2018, around 23% of trade volume in Ripple, the third largest cryptocurrency at the time of writing, occurred on Bithumb, a South Korean Exchange that is only open to South Korean citizens. During the same time, only 2% of Bitcoin and 3% of Ether volume occurred on Bithumb. Similarly, OKEX, an exchange registered in Hong Kong, accounted for 29% and 43% of two other major cryptocurrencies, Litecoin and Ethereum Classic respectively. At the same time, the exchange handled 12%, 12%, and 8% of Bitcoin, Ether, and Ripple respectively.[9] There is an even larger heterogeneity in market share of smaller currencies. This heterogeneity creates cross-sectional variation in the exposure of different cryptocurrencies to different investor bases formed on different trading locations. I exploit this variation to quantify the similarity in the investor bases of cryptocurrencies using a pairwise connectivity measure.

I.B Related Literature

This paper is related to three different strands of literature. First, my paper contributes to a fast-growing literature on cryptocurrencies. Previous studies in this literature have examined different aspects of the cryptocurrency market and the blockchain technology including broad economics of cryptocurrencies [Böhme et al. (2015), Harvey (2016), Raskin and Yermack (2016), Schilling and

[9] Figure IA1 shows this variation as a percentage of exchange volume for the 15 largest exchanges in my sample. For example, a large part of transactions on the two major South Korean exchanges, Bithumb and Coinone, belongs to Ripple.

Uhlig (2018), and Biais et al. (2018)], empirical asset pricing properties of cryptocurrencies [Liu and Tsyvinski (2018), Hu, Parlour, and Rajan (2018), and Liu, Tsyvinski, and Wu (2019)], market manipulation and illegal activities [Foley, Karlsen, and Putniņš (2018), Li, Shin, and Wang (2018), Gandal et al. (2018), and Griffin and Shams (2018)], consensus mechanism and adoption problem [Hinzen, John, and Saleh (2019a) and Hinzen, John, and Saleh (2019b)], mining activities and transaction costs [Easley, O'Hara, and Basu (2018), Cong, He, and Li (2018), and Pagnotta and Buraschi (2018)], and initial coin offerings [Kostovetsky and Benedetti (2018), Howell, Niessner, and Yermack (2018), Lee, Li, and Shin (2018), Li and Mann (2018), and Malinova and Park (2017)]. Most related to my paper, Gandal and Halaburda (2016), Sockin and Xiong (2018), and Cong, Li, and Wang (2018) relate the value of cryptocurrencies to the network externality effect of users' participation on the platform. Moreover, Kroeger and Sarkar (2017) and Makarov and Schoar (2018) show a sizable price difference across Bitcoin exchanges and attribute that to different investor bases and frictions to move capital. To the best of my knowledge, my paper is the first to examine the driving forces behind cryptocurrencies' comovement structure and document a strong amplification effect through the network externalities of users'/investors' decisions.

Second, this study is also closely related to the literature on demand-driven comovement. A broad literature in finance relates such comovement to common ownership and investment habitat [Barberis, Shleifer, and Wurgler (2005) and Bartram et al. (2015)], wealth effect [Kyle and Xiong (2001)], correlated liquidity shocks and trading behavior [Greenwood and Thesmar (2011) and Greenwood (2007)], and portfolio re-balancing [Fleming, Kirby, and Ostdiek (1998) and Kodres and Pritsker (2002)]. Most related to my paper, Anton and Polk (2014) "connect" stocks through mutual fund ownership and show that common ownership leads to excess comovement. My paper contributes to this literature by showing a new channel in which the comovement could be significantly amplified in a market where investment decisions have network externalities and assets have opaque fundamentals.

Finally, this paper relates to the literature on cross-listing that shows stocks tend to comove more with other stocks listed in the same location. Froot and Dabora (1999) and De Jong, Rosenthal, and Van Dijk (2009) show that dual-listed stocks tend to comove more with local stocks. Moreover, Chan, Hameed, and Lau (2003) and Brealey, Cooper, and Kaplanis (2009) show that changes in trading location significantly increases the comovement of a stock with the new market. Gagnon and Karolyi (2010) show that such an effect can lead to strong price deviations across markets.

This paper builds on these strands of literature and extend them by characterizing the return structure of a rich set of cross-listed assets in a fast-growing market. I present empirical evidence on a novel channel that significantly amplifies the effect of local demand shocks on prices and creates a strong comovement structure in cryptocurrency returns.

II. Testable Hypotheses, Empirical Framework, and Data

II.A Hypothesis Development

My empirical design is based on the premise that investors trade in a correlated manner across cryptocurrencies in their investable set. This assumption is informed by a broad literature on comovement as discussed above. Among other explanations, this pattern can arise in the cryptocurrency market due to country-specific sentiments or wealth and liquidity shocks that trigger trading across a wide set of cryptocurrencies. For example, a large swing in the price of Ripple can impact the wealth of South Korean crypto-traders who heavily trade Ripple, which can in turn spill over to other currencies listed in South Korea. As in Hasbrouck and Seppi (2001), such an effect could be reflected as commonalities in the order imbalance. My first hypothesis tests this assumption.

> *Hypothesis 1:* If investors have correlated demands across cryptocurrencies in their investable set, a common component should drive the order imbalance of cryptocurrencies listed on the same exchange, even after controlling for the imbalance of the same currency on other exchanges.

The next hypothesis describes how overlapping exposure to such correlated demand can drive cryptocurrencies' comovement. The novel features of cryptocurrencies can cause demand shocks to have substantial effects on returns. The underlying value of cryptocurrencies depends on the externality effect of users' participation on their platform [Sockin and Xiong (2018)]. Importantly, because using the features of a cryptocurrency ecosystem often necessitates developers and users to hold tokens, the user base of a cryptocurrency is inherently interwoven with its investor base. For example, to develop decentralized applications on the EOS platform, one of the largest cryptocurrency ecosystems, participants need to hold EOS tokens. As theoretically motivated by Cong, Li, and Wang (2018), agents decide how many cryptocurrencies to hold depending on both their user motive and investment motive. In this environment, if the market cannot distinguish between speculator and user demand, even pure speculative demand can have an amplified effect on prices. Moreover, opaque fundamentals of cryptocurrencies can further magnify the effect of demand shocks on prices.[10] Therefore, overlapping exposure to demand shocks can create a strong comovement in cryptocurrencies' returns.

> *Hypothesis 2:* If investors have correlated demand across cryptocurrencies, currencies that have similar investor bases should exhibit strong comovement in returns over and above what can be explained by the similarity in their characteristics. Additionally, exogenous variation in the investor base should cause changes in the comovement.

The next hypothesis examines the effect of technological features on return comovement. Cryptocurrencies with similar technical features might be exposed to similar fundamental shocks. For example, highly computationally intensive coins might comove because they are susceptible to similar news about computing advancements. Moreover, changes in people's perception of different technologies, such as the hashing algorithm or consensus mechanism, can cause a comovement in the price of cryptocurrencies that possess these features. Likewise, utility tokens might comove with other utility tokens and equity tokens with other equity tokens because they are exposed to similar legal shocks.[11]

> *Hypothesis 3:* If supply side factors and technological features are essential drivers of cryptocurrencies' underlying value, currencies with similar characteristics should exhibit a higher return comovement.

[10] Hirshleifer (2001) discusses that behavioral biases are strongest when information about fundamentals is more sparse. Kumar (2009) shows that individual investors are more prone to behavioral biases when stocks are harder to value. In a noisy dynamic rational expectations model, Cespa and Vives (2011) show that asset prices can deviate from their fundamental values when their fundamental is opaque.

[11] A full review of cryptocurrencies' main technological features used in this paper can be found in Table IAI.

The last hypothesis directly examines the role of the network externalities from the community of users and developers in amplifying demand shocks. Importantly, there is variation in the extent to which different cryptocurrencies derive value from the network effect of their user and developer community. For example, Ethereum provides an open source tool for users and developers to build applications on top of its platform using smart contracts. The underlying value of Ether, the main token used on the platform, largely depends on the participation of the community of users and developers. Similarly, the founder of the EOS, a platform for developing decentralized applications, describes holding the token as: "The token represents people's involvement in the *community*. At the end of the day, once you have a token, it is the *community* that is going to be responsible for taking this forward."[12] On the other hand, certain cryptocurrencies have an underlying business that depends less on the community for growth. For example, Ripple primarily expands its technology through contracted software engineers and grows revenue by signing contracts directly with large banks as their customers. Users, developers, and investors are not as interwoven for Ripple. If buying and selling pressures in the market are at least partially perceived to be from the potential users and developers, the price of cryptocurrencies that highly rely on user and developer community should react more to such demand pressures. This heterogeneity has a testable implication for the comovement of cryptocurrencies.

Hypothesis 4: If the externality effect of the user and developer community is an essential source of cryptocurrencies underlying value, "high-community-based" cryptocurrencies should exhibit a significantly larger comovement when exposed to similar demand shocks.

The next sections lay out the empirical design to test these hypotheses.

II.B Empirical Framework

II.B.1 Constructing the Connectivity Measure

To proxy for exposure to similar demand shocks, I define $Connectivity_{i,j,t}$ as a pairwise index that has a negative relationship with the Manhattan distance between the share of trading volume of currencies i and j across different exchanges:

[12]TV interview with BBC, December 15, 2017.

$$Connectivity_{i,j,t} = 1 - \frac{1}{2}\sum_{k=1}^{K} |p_{i,k,t} - p_{j,k,t}| \tag{1}$$

where $p_{i,k,t}$ is the share of trading volume of currency i in month t that occur on exchange k:

$$p_{i,k,t} = \frac{V_{i,k,t}}{\sum_{n=1}^{K} V_{i,n,t}} \tag{2}$$

and $V_{i,k,t}$ is the volume of currency i in month t on exchange k, and K is the total number of exchanges in my sample. The connectivity measure takes a value between zero and one where one indicates that the share of monthly volume of two currencies is perfectly identical across different exchanges, zero indicates they trade on entirely different locations, and a number in between indicates the extent of partial overlap. Given the trading environment described in Section I, a higher connectivity translates to exposure to more similar demand shocks.

My empirical design requires sufficient cross-sectional heterogeneity in this connectivity measure. As a first pass, the average cross-sectional standard deviation of the measure is 0.34 and the mean is 0.27. Figure 4 shows that the connectivity measure can be used to divide the cross-section of cryptocurrencies into multiple clusters of interconnected currencies and perhaps help tame the "Wild West" of cryptocurrencies. The figure illustrates a network graph where each node represents a cryptocurrency, and the edges represent the average monthly connectivity of each pair. A modularity analysis based on the strength of the connectivity between different nodes divides the cross-section of cryptocurrencies into several connected clusters as plotted with different colors. Cryptocurrencies in the same cluster are more likely to be traded on the same set of exchanges and be exposed to the same demand shocks.

There is also considerable heterogeneity within members of each cluster. Figure 5 zooms in on two clusters to examine this matter. For clarity, the graph only shows intra-cluster connections. As shown in both panels, there is a wide variation in the connectivity of members of a cluster with members of other clusters. This pattern arises because cryptocurrencies are cross-listed on several different exchanges, and as a result, there is wide variation in the connectivity even among currencies that share a trading location. My connectivity measure summarizes this complex structure into a pairwise index.

II.B.2 Relationship Between Return Comovement and Connectivity

The baseline analysis examines how variation in connectivity measure explains cross-sectional variation in pairwise comovements of cryptocurrencies. I measure comovement using within-month pairwise correlation of market model return residuals. Cryptocurrency exchanges are open 24/7 and do not have an opening or closing auction. To increase precision and avoid choosing an ad hoc closing time, market betas, return residuals, and pairwise correlations are calculated using rolling windows of 24-hour returns that move forward every hour. The calculation proceeds in several steps. First, volume-weighted cryptocurrency market return is calculated using the currencies' previous 24-hour trading volumes as the weights. Second, return residuals of the market model are estimated using recursive time-series regression of individual currencies' returns on the market returns. Finally, the comovement is calculated as the within-month correlation of the 24-hour return residuals for all pairwise combinations. I exclude currency pairs for which the pairwise returns are missing for more than 50% of the times in a given month.

I estimate a dyadic panel regression of return correlations on one-month lagged connectivity as reported in Equation (3). Previous studies including Koch, Ruenzi, and Starks (2016), Israelsen (2016), Anton and Polk (2014), and Kallberg and Pasquariello (2008) have used such a dyadic setting to examine comovement in stock returns and liquidity. I correct the standard errors with dyadic clustering approach of Cameron and Miller (2014), which takes into account that error terms for i-j pair might be correlated with those of any other currency pair that includes either i or j. All non-dummy variables are standardized by subtracting the mean and dividing by the standard deviation. Note that the unit of observation is currency pair per month.

$$Corr_{i,j,t} = \beta_0 + \beta_1 Connectivity_{i,j,t-1} + \beta^{Char} Similarity^{Char}_{i,j,t-1} + \delta_t + \varepsilon_{i,j,t} \tag{3}$$

$Similarity^{Char}_{i,j,t}$ is a vector of binary and continuous variables which directly control for similarity in characteristics. If both currencies have the same categorical characteristic, the relevant variable takes the value of one, and zero otherwise. For example, $Similarity^{CoinToken}$ takes the value of one if the currencies are either both coins or both tokens. Furthermore, for continuous characteristics, the similarity variable is defined as the negative of the absolute difference in the percentile rank-transformed value of that characteristic:

$$Similarity^{Char}_{i,j,t} = -|Pctl^{Char}_{i,t} - Pctl^{Char}_{j,t}| \tag{4}$$

where $Char$ is a continuous currency characteristic such as total trading volume. The next section describes the data for implementing the empirical design.

II.C Data

I construct a comprehensive dataset on cryptocurrencies by combining three types of data. The first type contains rich exchange-level trading and pricing data on a wide cross-section of cryptocurrencies from more than 70 trading platforms around the world. The second data source covers the main technical features of different cryptocurrencies collected from various online sources. The third dataset contains more than 12 million crypto-related comments from cryptocurrency-specific pages on the social news website Reddit. The result is a dataset containing rich information on a wide cross-section of cryptocurrencies. The number of currencies included in the sample varies from 50 at the beginning of the sample in January 2017 to more than 500 in June 2018.

II.C.1 Trading and Price Data

The main pricing data come from two of the most comprehensive data providers in the cryptocurrency space: *CoinAPI* and *Kaiko*. These vendors collect and aggregate intraday pricing and other trading data for thousands of cryptocurrencies listed on various trading platforms via the exchange APIs. To ensure the quality of the data, I merge these two datasets with data on the daily price and aggregate volume from *CoinMarketCap* and cross-check these datasets for coding errors and data problems. The Internet Appendix IA.A describes this procedure as well as the filters applied to the data.

The *CoinAPI* data contain a total of more than 18,000 pricing series for different currency pairs listed on more than 70 exchanges. The listed currencies are either denominated in fiat currencies (mainly US Dollar (USD), Euro, Chinese Yuan, Japanese Yen, and Korean Won), pegged digital currencies such as Tether, or other cryptocurrencies, primarily Bitcoin and Ether. For each currency pair on each exchange, I have access to minute-level open, close, high, and low prices as well as trading volume and number of trades.

The *Kaiko* data cover a smaller universe and contain approximately 4,500 pricing series from 26 exchanges. However, it includes the complete tick-level order book data with an indicator that shows whether a transaction is buyer- or seller-initiated. The data contain transaction IDs and timestamps of trades in Coordinated Universal Time (UTC), price and signed volume, and the bid/ask prices and sizes for the entire market depth.

Most cryptocurrencies trade on multiple exchanges, and prices can vary between these exchanges. I incorporate all available prices for each currency and calculate an aggregate return index. The procedure proceeds in several steps. First, prices quoted in fiat currencies other than USD are converted into USD based on exchange rates obtained from Bloomberg. Second, prices quoted in other cryptocurrencies such as Bitcoin are converted into USD using an aggregate volume weighted price index for Bitcoin using all Bitcoin-fiat prices across all exchanges. Next, I calculate the returns for each pricing series separately and exclude observations in the top and bottom 0.1 percentile. Finally, I calculate a weighted return index using all the return series weighted by the past 24-hour trading volume.[13] Table I reports the number of currencies with available returns in each month as well as the equal- and value-weighted daily returns for the currencies in my sample. The market experienced the highest returns in December 2017, with daily value-weighted returns of 4.7% and the lowest returns in March 2018 with daily -1.22% returns.

[13]Calculating weighted price indices first and then calculating the return index using those single price series produces virtually the same results.

II.C.2 Data on Technological Characteristics

Data on technological characteristics of cryptocurrencies are collected and cross-checked from various online sources. These characteristics include whether a cryptocurrency is a coin or token, hashing algorithm, consensus mechanism, being forks on the same blockchain, block time, passage of time from the genesis block, hash rate, token industry, token platform, and an identifier for whether a token is a utility or equity. Table IAI describes these characteristics and the source of the data.

II.C.3 Social Media Data

Finally, I collect 12 million cryptocurrency-specific comments from a social news website, *Reddit*. Most cryptocurrencies have their own *subreddit* page, where their members exchange information and discuss issues related to the currency. I first scrape the *subreddit* addresses of all cryptocurrencies from *CoinMarketCap* and then obtain all comments in these cryptocurrency-specific pages from *Reddit*.

III. Correlated Demand Across Cryptocurrencies

The empirical design of this paper rests on the initial hypothesis that investors have correlated demand across currencies listed on the same trading platform. This section examines this hypothesis by analyzing the drivers of order imbalance of different currencies listed on the same exchange.

III.A Constructing Proxies for Components of the Order Imbalance

To examine Hypothesis 1, I analyze the commonalities in the order imbalance of currencies listed on the same exchange. The order imbalance is calculated using the exchange reported buyer- and seller-initiated transactions. Transactions are designated as buyer- (seller-) initiated if the buy (sell) side of the trade placed the *market order*. The idea is that market makers or high-frequency trading bots who make money on both sides of the market make the book by placing *limit orders*. Liquidity demanders or privately informed impatient traders who have a desired directional position place market orders and initiate the trade. Therefore, the commonalities in order imbalance can reflect correlated directional trading across currencies listed on the same exchange.

The rich set of cross-listed cryptocurrencies allows me to examine this by disentangling the exchange-specific drivers of the order imbalance from the currency-specific component.[14] The analysis proceeds in several steps. First, I compute two measures of order imbalance for currency i on exchange k on day t. Similar to Chordia and Subrahmanyam (2004), I define $OIBVOL$ as the daily buyer-initiated volume less seller-initiated volume scaled by total volume, and $OIBNUM$ as the daily number of buyer-initiated less the number of seller-initiated transactions scaled by the total transaction:

$$OIBVOL_{i,k,t} = \frac{BuyVolume_{i,k,t} - SellVolume_{i,k,t}}{BuyVolume_{i,k,t} + SellVolume_{i,k,t}} \tag{5}$$

$$OIBNUM_{i,k,t} = \frac{BuyTransactions_{i,k,t} - SellTransactions_{i,k,t}}{BuyTransactions_{i,k,t} + SellTransactions_{i,k,t}} \tag{6}$$

Second, I construct three proxies for currency-specific, exchange-specific, and market-wide drivers of the order imbalance. The currency-specific component is the average order imbalance of currency i on all exchanges excluding exchange k ($\overline{OIB}^{Cur}_{i,k,t}$). The exchange-specific component is the average imbalance of all currencies on exchange k excluding currency i ($\overline{OIB}^{Exch}_{i,k,t}$). The market component is the average order imbalance of all currencies in the market excluding both currency i and all currencies on exchange k ($\overline{OIB}^{Mkt}_{i,k,t}$).

Finally, I estimate a panel regression of the order imbalance at currency-exchange level on the three proxies for the currency, exchange, and market components:

$$OIB_{i,k,t} = \beta_0 + \beta_1 \overline{OIB}^{Cur}_{i,k,t} + \beta_2 \overline{OIB}^{Exch}_{i,k,t} + \beta_3 \overline{OIB}^{Mkt}_{i,k,t} + \delta_{i,k} + \varepsilon_{i,k,t} \tag{7}$$

where $OIBi_{i,k,t}$ denotes the order imbalance for currency i on exchange k on day t (either $OIBVOL$

[14] A similar setting could be found in international markets with cross-listed stocks through dual-listing, American Depositary Receipt (ADR), or Global Registered Shares (GRS). First, the cross-listing is pervasive in the cryptocurrency market, and uniform order book data are accessible on a rich set of currencies across several trading platforms. Second, the cross-listed assets are identical and interchangeable, as opposed to dual-listed stocks that can behave differently because they are not necessarily convertible into each other and the arbitrage positions are risky.

or $OIBNUM$) and $\delta_{i,k}$ is currency-exchange fixed effect. The standard errors are two-way clustered by currency and exchange. I am interested in the coefficient β_2, which captures the investor-specific effect, controlling for the currency's own and the market-wide effect.[15] Note that all the three components on the right-hand side exclude the currency on the left-hand side.

III.B The Main Drivers of Order Imbalances

Panel A of Table II reports the results for $OIBNUM$, which uses the number of trades to calculate the order imbalance. All variables are standardized so that the coefficients are comparable. Column (1) shows that the order imbalance increases by 0.13 standard deviations when the currency's own imbalance on other exchanges increases by one standard deviation. Column (2) shows that the order imbalance increases by 0.31 standard deviations when the imbalance of other currencies on the same exchange increases by one standard deviation. Column (3) reports that the market-wide effect also has an explanatory power in a univariate regression. The multivariate specification in Column (4) shows that the exchange effect is economically and statistically significant even after controlling for currency's own imbalance and the aggregate market. The coefficient on the exchange component is 0.30, which is larger than the currency's own effect. The rest of the market becomes insignificant after controlling for the exchange and currency effects. Column (5) controls for the lead and lag currency-specific component and shows that the results cannot be explained by slower or faster reaction of traders on certain exchanges to the news. Panel B of Table II reports qualitatively similar findings for the volume-based order imbalance, $OIBVOL$.[16] The results in this section support Hypothesis 1, which states that investors have strong correlated demands across cryptocurrencies listed on the same exchange.

[15] An alternative approach would be to implement a principal component analysis to examine if a common component drives the order imbalances within an exchange. However, the regression setting here has the advantage of directly controlling for the currency's own and market-wide effects.

[16] Table IAII reports similar results for the sample of currencies that are listed on at least five exchanges and exchanges that operate at least five currencies. The results are qualitatively the same.

IV. Exposure to Overlapping Demand Shocks and Comovement in Returns

This section examines Hypotheses 2 and 3. I first test the relationship between comovement and the connectivity measure and other characteristics of cryptocurrencies in a dyadic setting as described in Section II.B. I then construct a "connected portfolio" return that summarizes the returns of all currencies connected to any given cryptocurrency to directly examine the implications of the connectivity for returns.

IV.A Connectivity and Return Comovement

This section estimates the relationship between lagged connectivity and comovement using the regression in Equation (3). Panel A of Table III reports the baseline results. For a one-standard-deviation higher connectivity, the next month's pairwise return correlation is higher by 0.19 standard deviations. The result is both statistically and economically large. The similarity in trading volume is also related to comovement. Cryptocurrencies with a one-standard-deviation higher similarity in volume comove by an additional 0.07 standard deviations. Moreover, having the same currency type is also positively related to the comovement. However, the effect of the connectivity measure is considerably stronger than other variables.

To test if the results hold for larger currencies, I estimate the same regression for the top 100 currencies based on the previous month's aggregate volume. Columns (4) to (6) show that the pattern is not just driven by small and less liquid cryptocurrencies and is similarly strong for larger currencies. Moreover, Panel A of Figure 6 shows that the comovement increases almost monotonically with connectivity. The figure sorts the cross-section of currencies into deciles of connectivity at each point in time, takes the average correlations for each decile, and reports the time-series mean for each decile.

Finally, Panel B of Table III documents similar results for comovement in order imbalance, where the comovement is calculated as pairwise correlation of 24-hour volume-based order imbalance as detailed in Section III. For a one-standard-deviation higher lagged connectivity, the correlation in order imbalance is 0.13 standard deviations higher.

One concern could be that there is a mechanical component in the relationship between connectivity and comovement. As described in Section II.C.1, I use volume-weighted returns across all exchanges to calculate comovement. Prices of currencies listed on the same exchange could be subject to the same exchange-wide noise or measurement errors. For example, noise in the exchange rate of Korean Won to USD can cause a spurious comovement in currencies heavily traded in Korea. One should note that the exchange-wide noise such as noise in FX rates are small relative to cryptocurrency price movements. Moreover, the results in section IV.D show that the effect increases with the time horizon, which is inconsistent with the noise explanation.

Nevertheless, I conduct a direct within-exchange analysis that is immune to exchange-wide noise. Because cryptocurrencies are often cross-listed on multiple exchanges, there is a wide variation in connectivity even among currencies that are listed on the same exchange. I first sort cryptocurrencies within each exchange based on their connectivity and then examine the pairwise correlations that are calculated only using prices on that exchange. Panel B of Figure 6 shows that more connected currencies comove more even when sorting within the exchange and using exchange-level prices, and the results are just as strong.[17]

IV.B Similarity in Characteristics and Return Comovement

This section examines how similarity in currency characteristics relates to return comovements. As discussed in Hypothesis 3, these characteristics can expose currencies to the same fundamental shocks. The similarity in these characteristics can also simultaneously determine trading location and lead to a spurious relationship between connectivity and comovement. Given different technical features, Table IV examines the effect of these characteristics separately for coins and tokens. Table IAI provides a detailed description of these characteristics.

The first three columns of Table IV, Panel A, examine digital coins. The dummy variable $Similarity^{HashAlgo}$ captures whether the pair has the same hashing algorithm, $Similarity^{ProofType}$

[17] Table IAIII shows that the magnitude of the effect is virtually unchanged relative to the baseline results.

captures similarity in consensus mechanisms, and $Similarity^{Fork}$ represents whether the two coins are forks on the same blockchain. Coins with similar proof type have 0.06 standard deviations higher returns correlations, which is marginally significant, but $Similarity^{HashAlgo}$ and $Similarity^{Fork}$ are statistically indistinguishable from zero. Moreover, the relationship between connectivity and return correlations holds strongly for the subset of coins and cannot be absorbed by the underlying technology.

The last three columns conduct a similar analysis for tokens. The variable $Similarity^{Platform}$ shows whether the pair is based on the same platform, $Similarity^{Industry}$ captures whether the pair belongs to the same industry, and $Similarity^{TokenType}$ determines whether the two tokens are both utility or equity. Column (6) shows that tokens in the same industry have 0.1 standard deviations higher return comovement. Moreover, the relationship between connectivity and comovement holds after controlling for these characteristics.

Finally, Panel B of Table IV controls for a more comprehensive set of characteristics available for 37 major coins, including similarity in block time, hash rate, and maturity (passage of time from the genesis block). First, most of these characteristics seem unrelated to the return comovement. Moreover, the connectivity measure is strongly related to the return comovement in this small sample of major coins. Overall, the results in this section suggest that although certain technological features are related to the returns, the effect of exposure to similar demand shocks bears the most substantial effect and technological, and other characteristics cannot absorb this effect.

IV.C Quantitative Magnitude

To gauge quantitative magnitude, this section examines (1) the percentage decile spread in realized correlations that can be explained by the connectivity and other characteristics and (2) the contribution of each variable to the cross-sectional R^2. To quantify the decile spread, I estimate the fitted values of the return correlations using Equation (3) with only connectivity index and time fixed effects included as explanatory variables. Then at each point in time, I sort currency pairs into deciles of the fitted values and calculate the average actual realized correlations for each decile. I then take the average values for each decile over time.

Panel A of Table V shows that fitted values using the connectivity alone generate a decile spread of 0.62 standard deviations in next months' pairwise correlations. This spread amounts to almost 18% of the realized spread. In comparison, fitted values estimated in the same fashion using the combination of all other characteristics including volume, hashing algorithm, consensus mechanism, token industry, and a host of other features explain only 8.5%.

Panel B of Table V shows the contribution of different sets of variables to the cross-sectional adjusted R^2. The first row in the left panel shows the average adjusted R^2 of monthly cross-sectional regression of return correlations on lagged connectivity. The first row of the right panel shows the incremental adjusted R^2 of connectivity when all other variables are already included. Both the univariate and the incremental R^2s show that the explanatory power of the connectivity measure is much larger than that of other variables. Given that the pairwise correlation of the market model residuals for individual cryptocurrencies is very noisy, the explanatory power of the connectivity measure appears quite strong. Overall, both the decile spread and the R^2 results show that the explanatory power of the connectivity measure alone is more than double the combination of all other characteristics and technological features.

IV.D Pricing Implications of the Connected Portfolio

This section directly investigates the return implications of the observed correlations by examining how returns of a currency respond to the returns of its "connected portfolio." To construct the connected portfolio return for currency i, I assign a weight to the returns of all currencies connected to i proportional to their connectivity to i and their trading volume as below.

$$R_{i,t}^{Con} = \sum_{j=1}^{N} w_{j,t} R_{j,t} \tag{8}$$

where N is total number of currencies and $w_{j,t}$ is defined as below:

$$w_{j,t} = \frac{Connectivity_{i,j,t-1} Volume_{j,t-1}}{\sum_{n=1}^{N} Connectivity_{i,n,t-1} Volume_{n,t-1}} \tag{9}$$

This portfolio return summarizes the returns of all currencies connected to currency i. The next two subsections examine the price movement of each cryptocurrency in relation to both contemporaneous and lagged returns of its connected portfolio.

IV.D.1 Contemporaneous Relationship

I first examine if cross-sectional variation in connected portfolio returns explains variation in the contemporaneous returns of individual currencies. Testing this relationship can be tricky because sorting the cross-section of currencies based on connected portfolio returns mechanically involves sorting on the currencies' own return as well. Lo and MacKinlay (1990) show how these two effects are intermingled in contrarian strategies. To avoid this mechanical relationship, when I compare connected portfolio returns of i and j, I exclude them in calculating the connected return of the other currency. I then assign a score of one to the currency with higher connected return and zero to the one with lower connected return. I conduct the same pairwise comparison for all currency pairs and sort currencies into 20 portfolios based on the average score. The average connected return for each portfolio is then assigned to each currency within the portfolio.

Panel A of Table VI reports cross-sectional Fama-MacBeth regressions of individual currency returns on the connected portfolio returns.[18] At the daily (weekly) horizon, one percent higher returns of the connected portfolio is associated with an economically large 32bp (57 bp) higher currency returns. The effect increases monotonically with the time horizon. Panel B of Table VI shows that the effect is even larger in 2018 with a daily (weekly) coefficient of 56bp (87bp). These results highlight the extent of covariation in currencies exposed to overlapping demand shocks.

IV.D.2 Investors' Underreaction and Cross-Predictability in Returns

The fact that the relationship between a currency's returns and its connected portfolio returns increases with the time horizon is consistent with investors underreacting to the information in the returns of connected currencies. In this section, I examine the efficiency of the market at incorporating the observed return structure into prices by testing if the return of connected portfolio can

[18] Since the results are estimated using overlapping returns with rolling windows that move forward every hour, the reported t-statistics are based on autocorrelation-consistent Newey-West standard errors.

predict individual currency returns. I create zero-cost portfolios with overlapping holding periods based on the lagged return of connected currencies. Every hour, I sort portfolios based on the past J days connected returns and hold them for the next K days. The portfolio is long currencies in the highest decile and short currencies in the lowest decile of lagged connected returns. Note that at any given hour, the portfolio holds $24K$ different overlapping portfolios constructed between the previous $24K - 1$ hours and the current hour. Therefore, the standard errors are adjusted using the Newey-West method based on the horizon of the overlapping returns.[19]

Table VII reports the average 24-hour returns of these zero-cost strategies. A strategy that buys currencies with past 24-hour winner connected currencies, shorts those with loser connected currencies, and holds them for the next 24 hours produces a before-fee return of 71bp per day with a t-statistic of 3.68. As illustrated by the time-trend of cumulative and compounded returns of the strategy in Figure 7, the effect is very persistent and does not diminish over time.

These results cannot be explained by non-synchronous trading for two reasons. First, my sample excludes illiquid currencies, and second, I calculate the holding period returns of the strategies from the first observed price after the sorting period.[20] This finding is consistent with a sizable inefficiency in cryptocurrency prices, potentially due to slow-moving capital or gradual diffusion of information.

V. Endogeneity of the Connectivity Measure

My connectivity measure picks up both listings of currencies on different trading locations (the extensive margin) and the popularity of the listed currencies among traders as reflected in the trading volume (the intensive margin). Both of these variables could be endogenous. Cryptocurrencies with similar unobservable fundamentals might comove together, and at the same time, they might be more likely to be co-listed on the same exchange and be traded by the same type of investors.

[19]For example, the reported standard errors of a strategy that holds k overlapping portfolios at any given time are adjusted using the Newey-West procedure with up to k lags.

[20]For example, if information up to 11:59:59 am on day t is used to sort portfolios, the holding period starts from the timestamp of the first trade observed after 12 pm.

I explore this issue in two settings. First, I examine if the comovement changes after cryptocurrencies are added to a new exchange. Second, I examine how a natural experiment that caused exogenous changes in the investor base of certain cryptocurrencies affected return comovement.

V.A New Exchanges Listing and Changes in Comovement

If the observed comovement pattern is due to the similarity in unobservable time-invariant characteristics of cryptocurrencies, the comovement of a newly-listed currency with the incumbent currencies should not be significantly affected after listing. I test this implication in a stacked-cohort difference-in-differences analysis for five major exchanges with accessible new listings information. These five exchanges are Binance, Bitfinex, Bittrex, Kraken, and Poloniex. I match each currency pair affected by the new listing with a control group of ten closest currency-pairs based on the Mahalanobis distance in the pairwise connectivity and the monthly trading volume of both currencies in the month before listing. I then estimate a regression of monthly pairwise return correlations on a treatment dummy variable interacted with month dummy variables, control variables, and the cohort-time and currency pair fixed effects:

$$\begin{aligned} Corr_{i,j,t} =& \beta_0 + \beta_1 Treated_{i,j} + \sum_{t=-1}^{3} \gamma_t Treated_{i,j} * M_t + \beta^{Char} Similarity^{Char}_{i,j,t-1} \\ &+ \sum_{t=-1}^{3} \gamma^{Char}_t Similarity^{Char}_{i,j,t-1} * M_t + \delta_{c,t} + \delta_{i,j} + \varepsilon_{i,j,t} \end{aligned} \tag{10}$$

where M_t is a dummy variable that takes the value of one in month t and zero otherwise, $\delta_{c,t}$ is the cohort-time fixed effect, and $\delta_{i,j}$ is the currency pair fixed effect. To avoid a mechanical change in the comovement, I exclude the returns of the added currency on the new exchange from my aggregate currency return index.

Table VIII reports the results. Time 0 represents the time of listing, and the baseline is $t = -2$. After the currency is listed, the comovement between the newly-listed currencies and the incumbent currencies increases significantly relative to the matched sample. Importantly, there is no pre-trend, and the comovement starts to increase in the month following the listing. Relative to the baseline, the difference in the comovements of the treated and control pairs increases by 0.30 standard deviations in three months. The relative connectivity measure of the treated pairs increases by 1.06 standard deviations during the same time. Thus, a one-standard-deviation increase in the connectivity leads to 0.28 standard deviations increase in the comovement. This effect is larger than the baseline results in Table III, suggesting that the extensive margin changes in the connectivity due to new listing have an explanatory power beyond that of the intensive margin.

V.B Shutdown of Chinese Exchanges and Exogenous Changes in Connectivity

Even though the comovement of cryptocurrencies changes significantly after cross-listing, the exchange listing is still endogenous. Cryptocurrencies might be listed on the same exchange due to the expectation about their future performance or certain technical features that turn out to be crucial only after the listing. I address this concern by constructing a plausibly exogenous instrument for changes in connectivity using a quasi-natural experiment that involves the shutdown of major Chinese exchanges by the Chinese government.

A series of crypto-related events occurred in China in early September 2017. First, the Chinese government banned raising funds through ICOs in September 4, 2017. Later on, the People's Bank of China (PBoC) ordered executives of cryptocurrency trading platforms to stop registering new clients and release plans for ceasing all operations in the near future.[21] Following these events, the trading activity of Chinese exchanges in my data gradually phased out by October 2017. At the time, nearly 80 currencies were cross-listed on these exchanges.

This event creates exogenous changes in the connectivity. Suppose that half the trading of currencies i and j occurs on the same Chinese exchange and the other half on two different exchanges outside China. While the two currencies are partially connected, they will have completely different investor bases after the shutdown of the Chinese exchange. My Hypothesis 2 predicts that the comovement of the two currencies on the exchanges outside China drops after the shutdown.[22]

[21] More information can be found here: https://www.loc.gov/law/help/cryptocurrency/china.php

[22] This example describes a negative shock to the connectivity. However, the event can cause a positive shock in a similar fashion. For example, suppose that i and j trade on the same exchange outside China, but half of i's trading occurs on a Chinese exchange. After the shutdown, the two currencies would be exposed to exactly the same demand

I use the pre-shutdown contribution of the Chinese exchanges to the connectivity of the affected pairs as an instrument for the actual changes after shutdown. The pre-shutdown contribution is calculated as:

$$Connectivity_{i,j}^{Chinese} = Connectivity_{i,j} - Connectivity_{i,j}^{NonChinese} \tag{11}$$

where $Connectivity_{i,j}^{NonChinese}$ is the connectivity measure constructed exactly as described in Section II.B.1 but only including the non-Chinese exchanges. $Connectivity_{i,j}^{Chinese}$ represents the negative of the implied changes in connectivity of i and j and can be used as an instrument for actual changes in the two-stage least-squares regression of the form:

$$First\ Stage: \quad \Delta Connectivity_{i,j} = \gamma_0 + \gamma_1 Connectivity_{i,j}^{Chinese} + \delta_c + \varepsilon_{i,j} \tag{12}$$

$$Second\ Stage: \quad \Delta Corr_{i,j} = \beta_0 + \beta_1 \Delta \widehat{Connectivity}_{i,j} + \delta_c + \varepsilon_{i,j} \tag{13}$$

where $\Delta Corr_{i,j}$ is the change in the average within-month return correlations from three months pre-treatment to three months post-treatment, $\Delta \widehat{Connectivity}_{i,j}$ represents the fitted values from the first stage regression, and δ_c is the matching cohort fixed effect. For each treated cryptocurrency pair, I match a control group by finding the ten closest currency-pairs based on the Mahalanobis distance in the pairwise connectivity and the monthly trading volume of both currencies the month before the shutdown.

Table IX reports the results. The first stage Wald F-statistic is 21.1, indicating a valid instrument. The second stage results show that a one-standard-deviation exogenous change in the connectivity causes a large 0.17 standard deviations change in return correlations. Figure 8 reports how the comovement of the currency pairs with negative connectivity shocks changes relative to the matched sample over time.[23] The comovements in the treatment and control groups are statistically indistinguishable before the shutdown. However, the comovement in the treated pairs significantly decreases after the shock. These results suggest that exogenous variations in the in-

shocks, while their investor base only partially overlapped before. For simplicity, the narrative explains a negative shock, but the results include both negative and positive changes.

[23]Figure IA3 reports the results for pairs with positive connectivity shocks.

vestor base of cryptocurrencies economically and statistically affect their return comovements.

Overall, the results in this section show that changes in the investor base of different currencies due to exchange-listing lead to changes in comovement, and plausibly exogenous shocks to the investor base cause sizable changes in the comovement. These findings highlight the substantial effect of investor demand on cryptocurrencies return structure.

VI. The Network Externality Effect of the User and Developer Community

This section examines Hypothesis 4 which predicts that the reliance of cryptocurrencies on the network externality, which comes from participation of user and developer community, can amplify the effect of demand shocks on prices. In order to categorize the cross-section of cryptocurrencies based on their reliance on this network effect, I collect currency-specific comments from a social news website, Reddit, and classify them as to whether they relate the value of the currency to the network effect using two methods. The first method relies on features extracted using a machine learning technique, and the second method uses the frequency of the most important feature extracted from the first method, "community," to classify the comments. I then test if exposure to similar investor base translates into a different comovement pattern for currencies that rely more on this effect.

VI.A Quantifying the Reliance on the Network Effect

The network effect from community building and user adoption is considered a primary source of cryptocurrencies' underlying value, not only in the academic circle, but also in the cryptocurrency community. There is abundant evidence in cryptocurrencies online forums and social media pages where investors and supporters relate the underlying value of a cryptocurrency to the network effect from the growing community. However, due to different nature of different cryptocurrencies, there is variation in the extent to which members of a given currency are concerned with this aspect. For example, those in Ethereum community talk more frequently about community building and the network effect than an average currency. I use this variation in the investors' and supporters' belief about different currencies to test the pricing implications of the network effect.

To quantify the reliance of different currencies on the network effect, I analyze 12 million currency-specific comments from Reddit, where members of currency-specific forums exchange information and discuss issues related to a specific currency. I use two methods to categorize the comments. The first method follows a standard machine learning technique and proceeds in several steps. First, a random training subsample of 10,000 comments were manually labeled as whether a given comment relate the underlying value of a cryptocurrency to concepts such as network effect, user participation, and community building. Second, the text of the comments was processed using a natural language processing package and converted into a feature matrix that shows the frequency of each word in each comment.[24] Third, the random forest method was applied using the training sample to determine the features that are important in distinguishing the comments that are labeled as network based.

Figure 9 shows these features and their relative importance. First, words containing "commun," such as "community" and "communities" are the most important feature in categorizing the comments. Also, both simple words containing "adopt", "user", "grow", "develop", "network", and more sophisticated words such as "ecosystem" and "metcalflaw" are among important words in the network-related comments, and the former has a higher relative importance. The rest of 12 million comments were labeled based on these features obtained from the trained model. The second method uses a straightforward measure motivated by the features extracted from the first method. It quantifies the frequency of the most important feature extracted from the first method, "community," to classify the comments. Cryptocurrencies are labeled as "high-community-based" if their average monthly percentage of network-based comments is above the median and "low-community-based" otherwise.

Several observations in the data support that these measures are reliable in capturing how much a cryptocurrency derives value from the network effect. First, there is a highly persistent compo-

[24]The text analysis is done using the Natural Language ToolKit (NLTK), by tokenizing the text into distinct words, removing stopwords ("a," "and," "but," "the," etc.), removing numbers, special characters, currencies names and tickers, and finally stemming the words using PorterStemmer.

nent in these measures and members of certain cryptocurrency forums persistently discuss more issues related to the network effect and community building. Panel A of Table X shows that if the percentage usage of *community* is higher by 1% for a cryptocurrency in the cross-section, the next month's percentage is likely to be higher by 0.63%. The effect is significant up to three lags and is consistent with persistent information content in the measure. Second, these measures are highly consistent with common knowledge about major cryptocurrencies. For example, Panel A of Figure 10 shows that members in the Ethereum page persistently discuss community-related issues to a noticeably higher extent than those of Ripple. The average measure is 3.4% for Ether, 1.6% for Bitcoin, and 0.8% for Ripple. Moreover, certain large cryptocurrencies known as "platform tokens" rely more on their community of developers and users than non-platform tokens. Due to the minimal number of these platform tokens, such as Ether and EOS, it is not feasible to divide the large cross-section of currencies merely based on this feature. However, Panel B of Table X shows that for nine out of ten large cryptocurrencies, being labeled as high- or low-community coincides with being a platform token or not. Therefore, the information that the measure captures is highly consistent with actual distinctive characteristics of these major cryptocurrencies.

Finally, Panel B of Figure 10 shows that consistent with the endogenous user adoption effect in Cong, Li, and Wang (2018), more community-based currencies show significantly higher volatility. The idea is that shocks to the fundamentals are amplified through endogenous adoption or attrition of participants in the ecosystem of cryptocurrencies that derive more value from the externalities of users' and developers' participation. The relationship is statistically significant and cannot be explained by characteristics such as trading volume, technical features, or the number of comments. Overall, these observations suggest that the measures developed here are reliable proxies for how community-based a cryptocurrency is. The next section examines the implications of these measures for return comovements.

VI.B The Amplification Effect of the Network Externalities

Hypothesis 4 predicts that exposure to overlapping demand shocks should increase the comovement of community-based currencies by a larger extent. To examine this effect, I use the measures developed in the previous section and test if the interaction of a high-community-based dummy variable with connectivity is related to return correlations:

$$\begin{aligned} Corr_{i,j,t} =& \beta_0 + \beta_1 Connectivity_{i,j,t-1} + \beta_2 HI\_COMM_{i,j} \\ &+ \beta_3 Connectivity_{i,j,t-1} * HI\_COMM_{i,j} + \delta_t + \varepsilon_{i,j,t} \end{aligned} \tag{14}$$

where $HI\_COMM_{i,j}$ is a dummy variable that takes the value of one if both currencies i and j are high-community-based and zero if both are low-community-based.

Table XI reports the results. The interaction between $Connectivity_{i,j,t-1}$ and $HI\_COMM_{i,j}$ is economically and statistically significant using both measures. Panel A reports the results for the measure based on the random forest model. Columns (1) to (3) suggest that the exposure to overlapping demand shocks increases the comovement of high-community-based cryptocurrencies by 43% to 51% more relative to low-community-based currencies. Column (4) shows that the results hold when only currency pairs that are both coins or both tokens are included in the analysis. Panel B shows the results for the second measure and suggest a similar pattern as Panel A. Columns (1) to (4) suggest that the exposure to overlapping demand shocks increases the comovement of high-community-based cryptocurrencies by 37% to 43% more relative to low-community-based currencies. These results highlight the novel economic channel that the effect of demand shocks on prices can be significantly amplified when investment decisions have network externalities.

VII. Conclusion

Investment in cryptocurrencies has soared and numerous technology startups have been funded using blockchain technology. Accordingly, understanding the drivers of cryptocurrency prices can have important investment and capital allocation implications. In particular, the interwoven nature of users and investors in this market can significantly amplify the effect of investor/user demand on prices, especially given the opaque fundamentals of these assets.

Consistent with this intuition, this paper finds that overlapping exposure to demand shocks, proxied by a pairwise "connectivity" measure based on trading locations, is a first order driver of the comovement structure in the cryptocurrency market. The explanatory power of the connectivity measure for return comovements is more than double that of all technological features and other characteristics combined. The results from new exchange listings and a quasi-natural experiment drawn from the shutdown of Chinese exchanges indicate that the findings are not driven by the endogenous sorting of currencies into different exchanges. Moreover, consistent with the notion that the entwined nature of the roles of users and investors can amplify the impact of demand shocks, I find that the effect is 40 to 50% larger for currencies that rely more on their user community.

Overall, the unique characteristics of the cryptocurrency market suggest that understanding the demand side of the market can be a starting point to assessing the valuation and price movements of these digital assets. Due to novel features of the market, the demand from users and developers may correlate with that of investors and speculators, implying that even pure speculative demand can have a substantial effect on prices. Further research is needed to disentangle the roles of users and investors in this market and understand the complex interplay between the two groups.

References

Anton, M. and C. Polk (2014). Connected stocks. *The Journal of Finance 69*(3), 1099–1127.

Barberis, N., A. Shleifer, and J. Wurgler (2005). Comovement. *Journal of financial economics 75*(2), 283–317.

Bartram, S. M., J. M. Griffin, T.-H. Lim, and D. T. Ng (2015). How important are foreign ownership linkages for international stock returns? *The Review of Financial Studies 28*(11), 3036–3072.

Biais, B., C. Bisiere, M. Bouvard, C. Casamatta, and A. J. Menkveld (2018). Equilibrium bitcoin pricing. *Available at SSRN*.

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

Brealey, R. A., I. A. Cooper, and E. Kaplanis (2009). Excess comovement in international equity markets: Evidence from cross-border mergers. *The Review of Financial Studies 23*(4), 1718–1740.

Cameron, A. C. and D. L. Miller (2014). Robust inference for dyadic data. *Unpublished manuscript, University of California-Davis*.

Cespa, G. and X. Vives (2011). Dynamic trading and asset prices: Keynes vs. hayek. *The Review of Economic Studies 79*(2), 539–580.

Chan, K., A. Hameed, and S. T. Lau (2003). What if trading location is different from business location? evidence from the jardine group. *The Journal of finance 58*(3), 1221–1246.

Chordia, T. and A. Subrahmanyam (2004). Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics 72*(3), 485–518.

Cong, L. W., Z. He, and J. Li (2018). Decentralized mining in centralized pools.

Cong, L. W., Y. Li, and N. Wang (2018). Tokenomics: Dynamic adoption and valuation.

De Jong, A., L. Rosenthal, and M. A. Van Dijk (2009). The risk and return of arbitrage in dual-listed companies. *Review of Finance 13*(3), 495–520.

Easley, D., M. O'Hara, and S. Basu (2018). From mining to markets: The evolution of bitcoin transaction fees. *Available at SSRN 3055380*.

Fleming, J., C. Kirby, and B. Ostdiek (1998). Information and volatility linkages in the stock, bond, and money markets1. *Journal of financial economics 49*(1), 111–137.

Foley, S., J. Karlsen, and T. J. Putniņš (2018). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?

Froot, K. A. and E. M. Dabora (1999). How are stock prices affected by the location of trade? *Journal of financial economics 53*(2), 189–216.

Gagnon, L. and G. A. Karolyi (2010). Multi-market trading and arbitrage. *Journal of Financial Economics 97*(1), 53–80.

Gandal, N. and H. Halaburda (2016). Can we predict the winner in a market with network effects? competition in cryptocurrency market. *Games 7*(3), 16.

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

Garcia, D., C. J. Tessone, P. Mavrodiev, and N. Perony (2014). The digital traces of bubbles: feedback cycles between socio-economic signals in the bitcoin economy. *Journal of the Royal Society Interface 11*(99), 20140623.

Greenwood, R. (2007). Excess comovement of stock returns: Evidence from cross-sectional variation in nikkei 225 weights. *The Review of Financial Studies 21*(3), 1153–1186.

Greenwood, R. and D. Thesmar (2011). Stock price fragility. *Journal of Financial Economics 102*(3), 471–490.

Griffin, J. M. and A. Shams (2018). Is bitcoin really un-tethered?

Harvey, C. (2016). Cryptofinance.

Hasbrouck, J. and D. J. Seppi (2001). Common factors in prices, order flows, and liquidity. *Journal of financial Economics 59*(3), 383–411.

Hayes, A. S. (2017). Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics 34*(7), 1308–1321.

Hinzen, F. J., K. John, and F. Saleh (2019a). Bitcoin's fatal flaw: The limited adoption problem. *SSRN (3334262)*.

Hinzen, F. J., K. John, and F. Saleh (2019b). Proof-of-work's limited adoption problem. *NYU Stern School of Business*.

Hirshleifer, D. (2001). Investor psychology and asset pricing. *The Journal of Finance 56*(4), 1533–1597.

Howell, S. T., M. Niessner, and D. Yermack (2018). Initial coin offerings: Financing growth with cryptocurrency token sales. Technical report, National Bureau of Economic Research.

Hu, A., C. A. Parlour, and U. Rajan (2018). Cryptocurrencies: Stylized facts on a new investible instrument.

Israelsen, R. D. (2016). Does common analyst coverage explain excess comovement? *Journal of Financial and Quantitative Analysis 51*(4), 1193–1229.

Kallberg, J. and P. Pasquariello (2008). Time-series and cross-sectional excess comovement in stock indexes. *Journal of empirical finance 15*(3), 481–502.

Koch, A., S. Ruenzi, and L. Starks (2016). Commonality in liquidity: a demand-side explanation. *The Review of Financial Studies 29*(8), 1943–1974.

Kodres, L. E. and M. Pritsker (2002). A rational expectations model of financial contagion. *The journal of finance 57*(2), 769–799.

Kostovetsky, L. and H. Benedetti (2018). Digital tulips? returns to investors in initial coin offerings.

Kroeger, A. and A. Sarkar (2017). Law of one bitcoin price? *Federal Reserve Bank of Philadelphia*.

Kumar, A. (2009). Hard-to-value stocks, behavioral biases, and informed trading. *Journal of Financial and Quantitative Analysis 44*(6), 1375–1401.

Kyle, A. S. and W. Xiong (2001). Contagion as a wealth effect. *The Journal of Finance 56*(4), 1401–1440.

Lee, J., T. Li, and D. Shin (2018). The wisdom of crowds and information cascades in fintech: Evidence from initial coin offerings.

Li, J. and W. Mann (2018). Initial coin offering and platform building.

Li, T., D. Shin, and B. Wang (2018). Cryptocurrency pump-and-dump schemes. *Available at SSRN*.

Liu, Y. and A. Tsyvinski (2018). Risks and returns of cryptocurrency. Technical report, National Bureau of Economic Research.

Liu, Y., A. Tsyvinski, and X. Wu (2019). Common risk factors in cryptocurrency. Technical report, National Bureau of Economic Research.

Lo, A. W. and A. C. MacKinlay (1990). When are contrarian profits due to stock market overreaction? *The review of financial studies 3*(2), 175–205.

Makarov, I. and A. Schoar (2018). Trading and arbitrage in cryptocurrency markets.

Malinova, K. and A. Park (2017). Market design with blockchain technology.

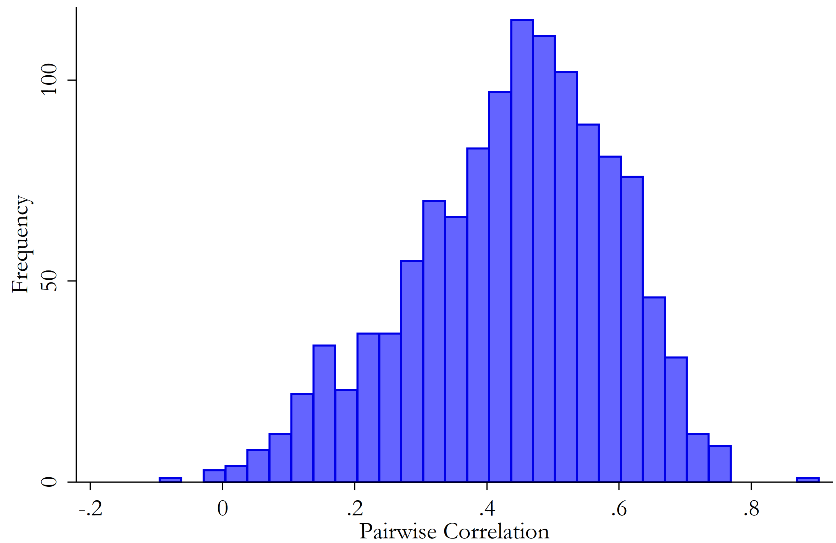
Pagnotta, E. and A. Buraschi (2018). An equilibrium valuation of bitcoin and decentralized network assets.

Raskin, M. and D. Yermack (2016). Digital currencies, decentralized ledgers, and the future of central banking. Technical report, National Bureau of Economic Research.

Schilling, L. and H. Uhlig (2018). Some simple bitcoin economics. Technical report, National Bureau of Economic Research.

Sockin, M. and W. Xiong (2018). A model of cryptocurrencies. *Unpublished manuscript, Princeton University*.

A. Distribution of Pairwise Correlations



B. Persistence of of Pairwise Correlations

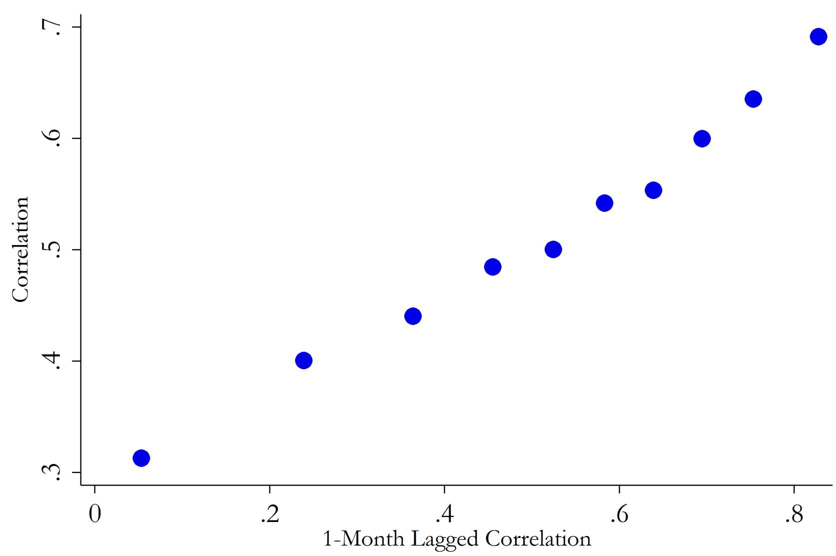


Figure 1. Distribution and Persistence of of Pairwise Correlations. This figure shows the persistence and distribution of the pairwise correlation of the largest 50 cryptocurrencies. The largest 50 cryptocurrencies are selected based on their market cap reported by *CoinMarketCap* as of January 1, 2017. Panel A shows the correlation of daily returns from January 1, 2017 to June 30, 2018. Panel B shows monthly correlation of daily returns for pairs sorted into deciles based on their correlation in previous month. the sample is from January 1, 2017 to June 30, 2018.

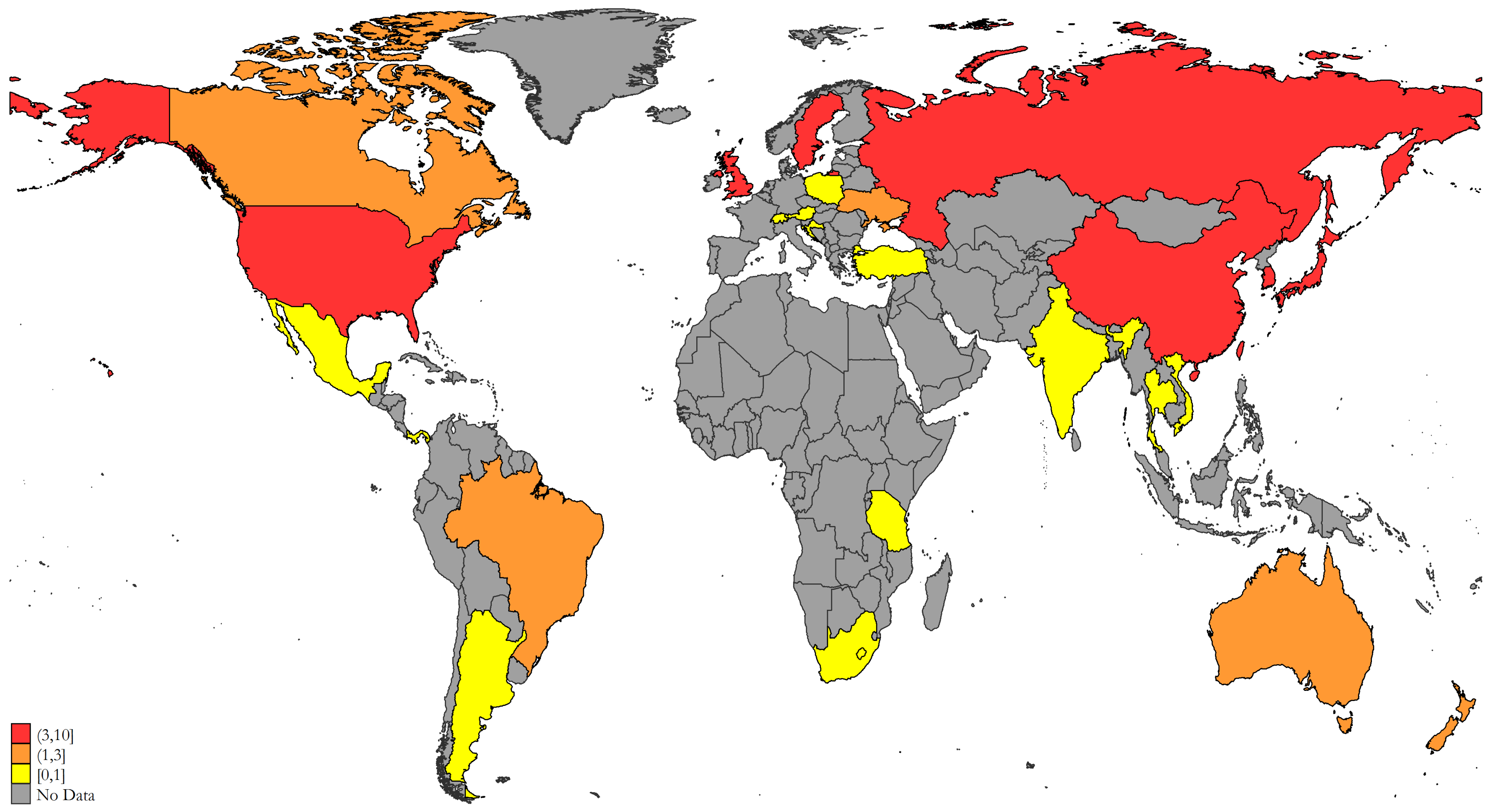


Figure 2. Geographical Distribution of Cryptocurrency Exchanges in the Data. This figure shows the geographical distribution of all cryptocurrency exchanges available in *CoinAPI* and *Kaiko* data from January 1, 2017 to June 30, 2018. Exchange locations are hand-collected from the exchanges' websites and, if not reported, from other public online sources. Countries with one exchange are in yellow, those with two or three exchanges are in orange, and those with more than three exchanges are in red. Countries that do not have an exchange in the data are in gray.

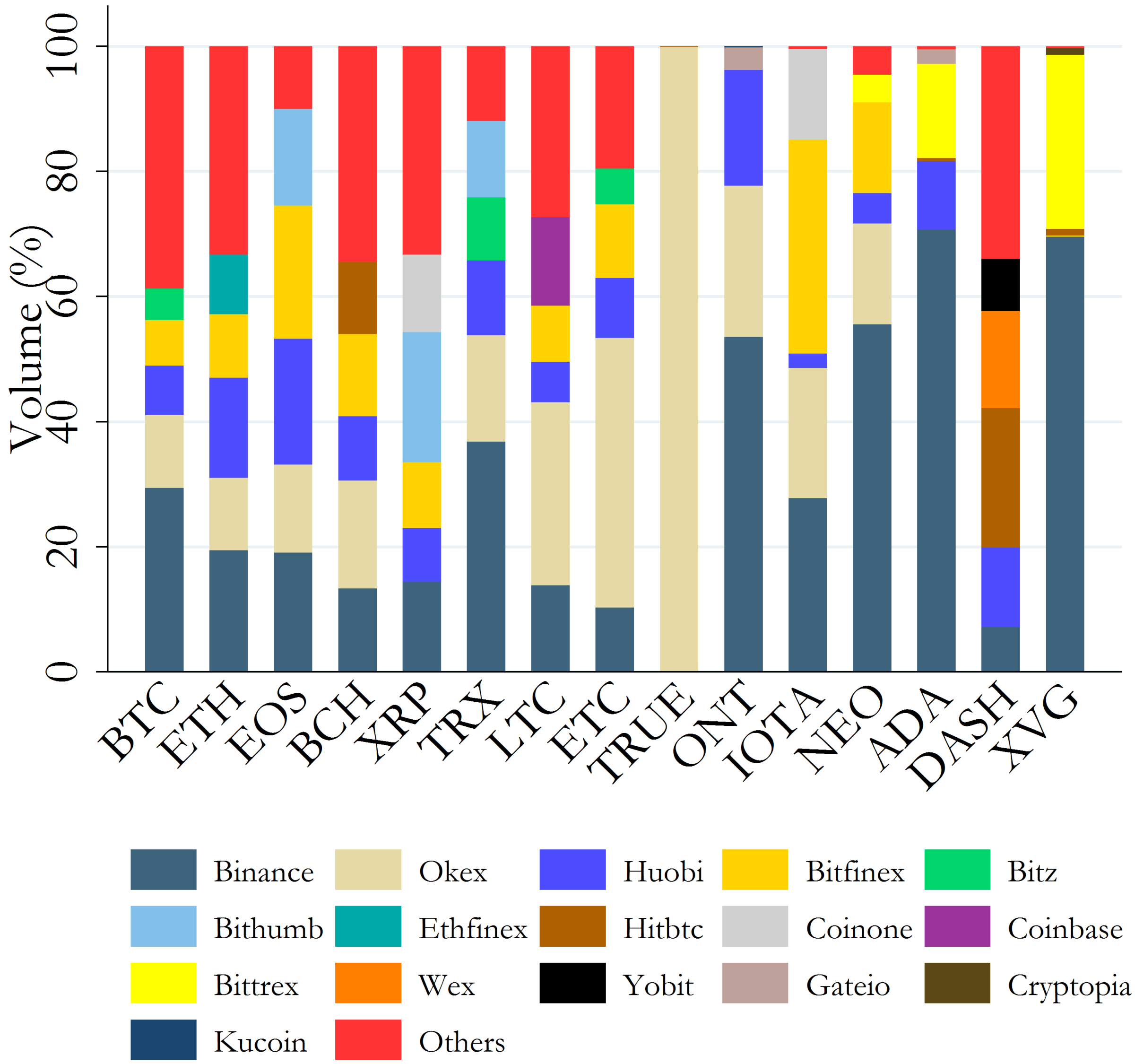


Figure 3. Variation in Percentage of Trading Volume of Major Cryptocurrencies on Different Exchanges. This figure shows the variation in trading volume of major cryptocurrencies across different exchanges available in *CoinAPI* and *Kaiko* data in the second quarter of 2018. Major cryptocurrencies are selected as the top 15 currencies in terms of dollar trading volume in this period. For each cryptocurrency, the figure plots the volume across different exchanges as the percentage of total volume of that currency. The currencies are sorted based on trading volume, where the volume decreases from left to right. Pegged cryptocurrencies are excluded.

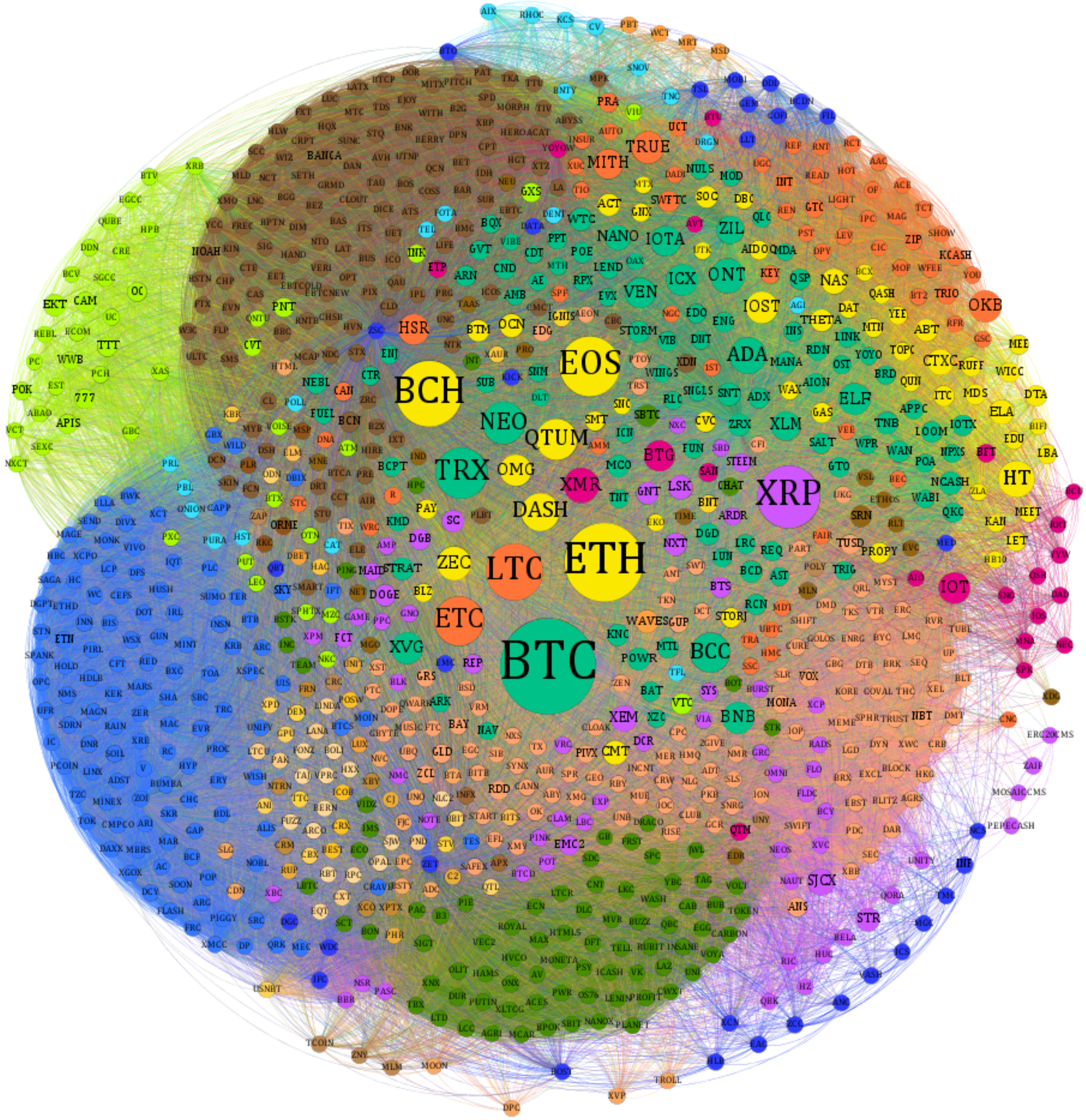
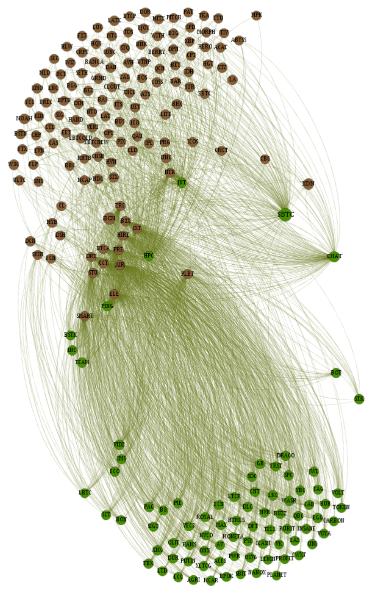


Figure 4. Different Clusters of Cryptocurrencies Based on the "Connectivity" Measure. This figure shows the network of cryptocurrencies. Each node represents a cryptocurrency, where the relationship between each pair of currencies is defined as the average monthly connectivity. The pairwise connectivity measure is calculated each month as described in Section II.B.1. The graph plots all currencies in the sample between January 1, 2017 to June 30, 2018. Different colors represent different clusters derived from a modularity analysis of the network structure. Node size is weighted by the average monthly dollar volume of each cryptocurrency.

A. Original Layout



B. Reorganized Layout

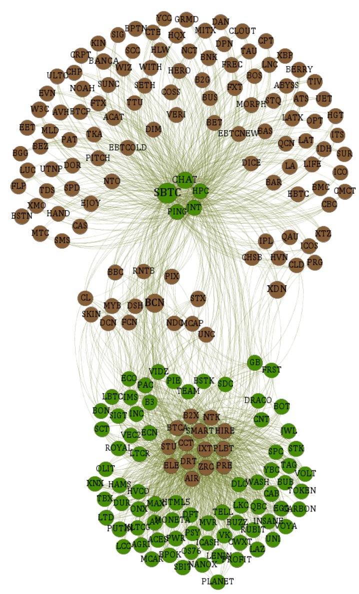
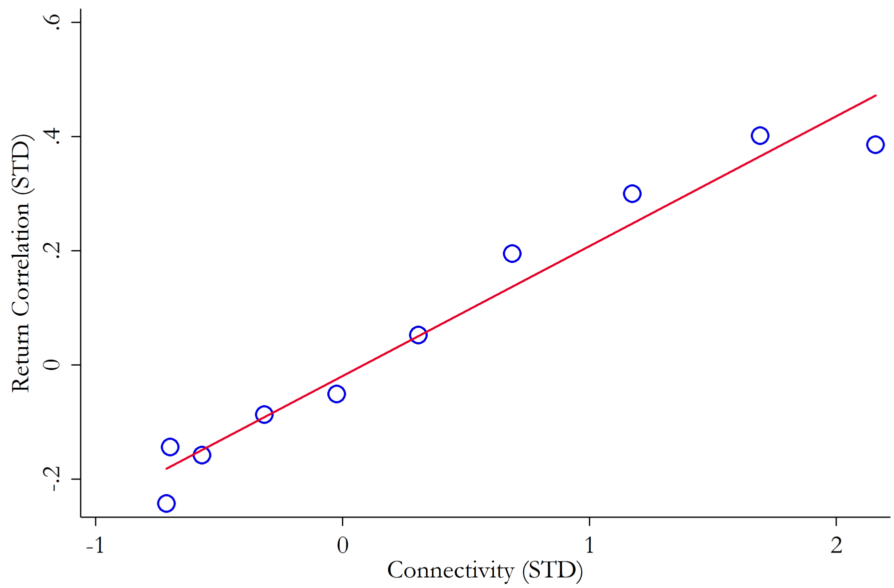


Figure 5. Example of Within-Cluster Heterogeneity in Connectivity with Members of Other Clusters. This figure shows the examples of within-cluster heterogeneity in the connectivity measure for two clusters in Figure 4. For clarity, the inter-cluster edges are removed, and the graph only shows the intra-cluster connections. Panel A shows the same layout as in Figure 4, and Panel B shows the reorganized layout where the distance between cryptocurrencies is optimized using intra-cluster relations based on the currencies in these two clusters.

A. Sorted Across All Exchanges



B. Sorted Within Exchanges

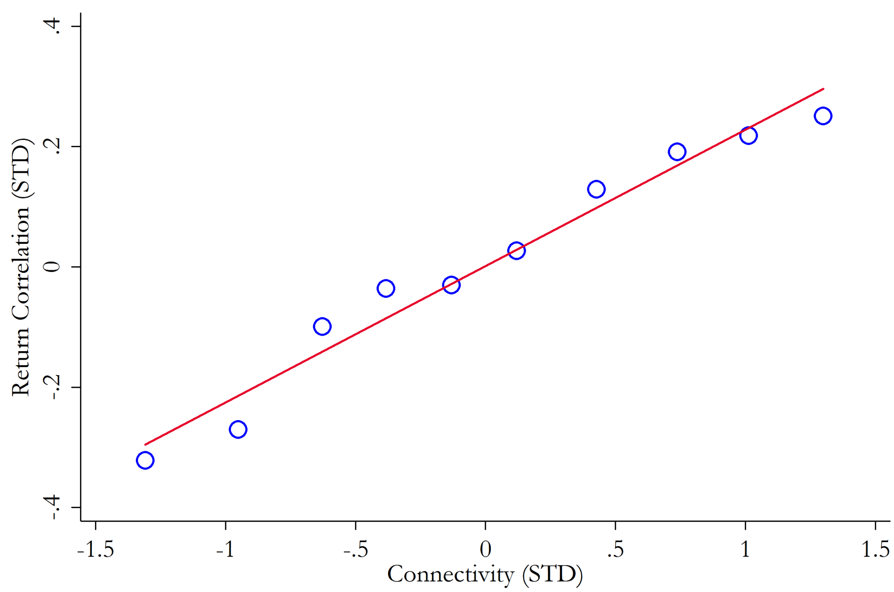
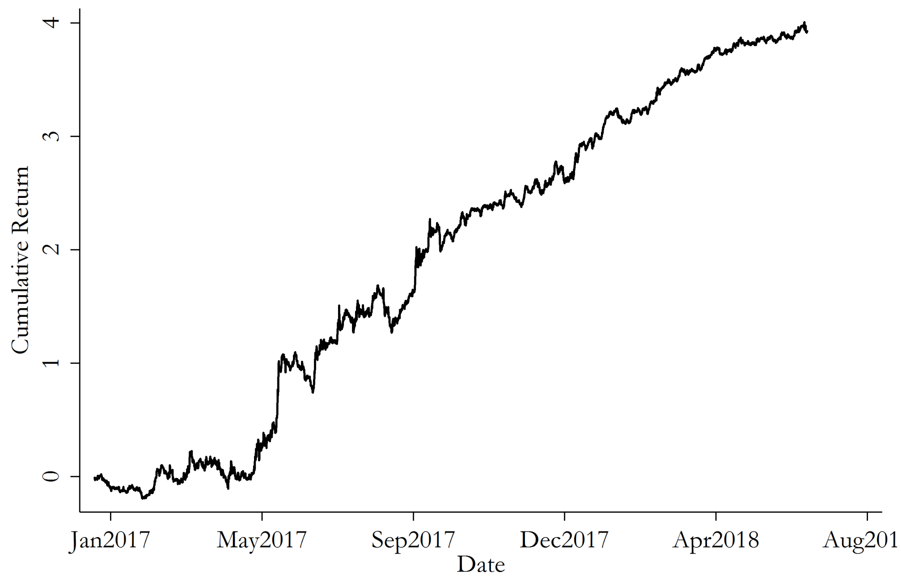


Figure 6. The Relationship between Comovement and Lagged Connectivity. This figure shows the relationship between one-month lagged connectivity and standardized correlation of the market model return residuals. The connectivity measure is calculated as described in Section II.B.1 and the returns correlations as in Section II.B.2. Each month, cryptocurrencies are sorted into deciles of lagged connectivity. Then the standardized return correlations are averaged for each decile in each month and then over time. Panel A sorts the pooled cross-section of all cryptocurrencies into deciles of connectivity and calculates the return correlations using an aggregate return index across all exchanges. Panel B sorts cryptocurrencies within each exchange first and then reports return correlations that are calculated only using prices on that exchange. The sample period is from January 1, 2017 to June 30, 2018.

A. Cumulative Returns



B. Compounded Returns

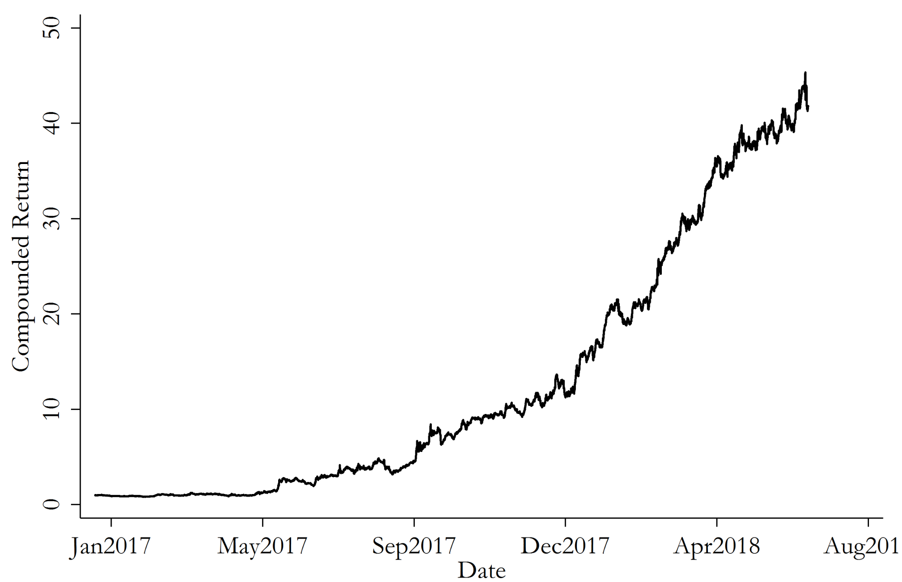


Figure 7. Returns on Zero-Cost Trading Strategy Based on the Connected Portfolio Returns. This figure shows the cumulative and compounded before-fee returns of a strategy that buys cryptocurrencies in the tenth decile of past 24-hour connected portfolio returns, sells those in the first decile, and holds the portfolio for the next 24 hours. The portfolio is rebalanced every hour, and at each point in time, the strategy holds 24 overlapping portfolios constructed between the previous 23 hours and the current hour. The connected portfolio return is calculated as in Section IV.D. The sample period is from January 1, 2017 to June 30, 2018.

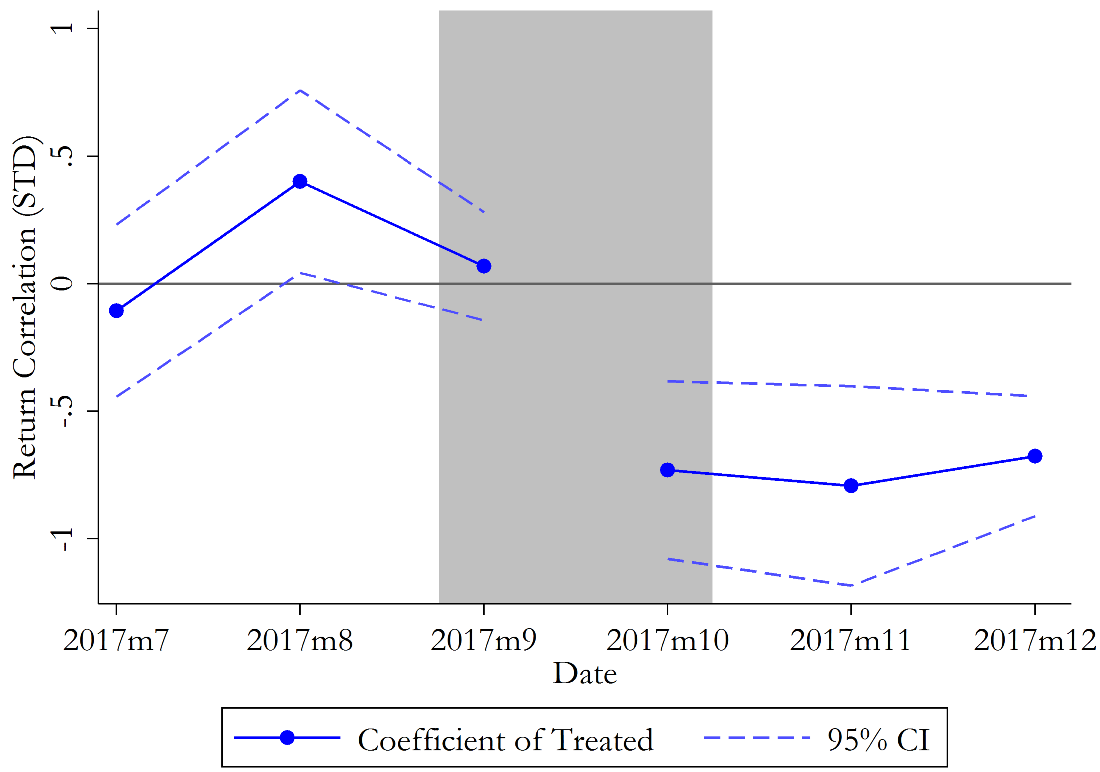


Figure 8. Shutdown of Chinese Exchanges and Changes in Return Correlations. This figure shows relative changes in return correlations of cryptocurrency pairs whose connectivity was negatively affected by the shutdown of Chinese exchanges. For each treated cryptocurrency pair, a control group is matched by finding the ten closest pairs based on the Mahalanobis distance in the pairwise connectivity and the monthly trading volume of both currencies the month before the shutdown. This figure shows the coefficient estimates of the standardized pairwise correlations on a treatment dummy controlling for similarity in volume, number of exchanges, and currency type. The blue dashed lines show the 95% confidence interval. The standard errors for estimation of the confidence intervals are clustered at the dyadic level.

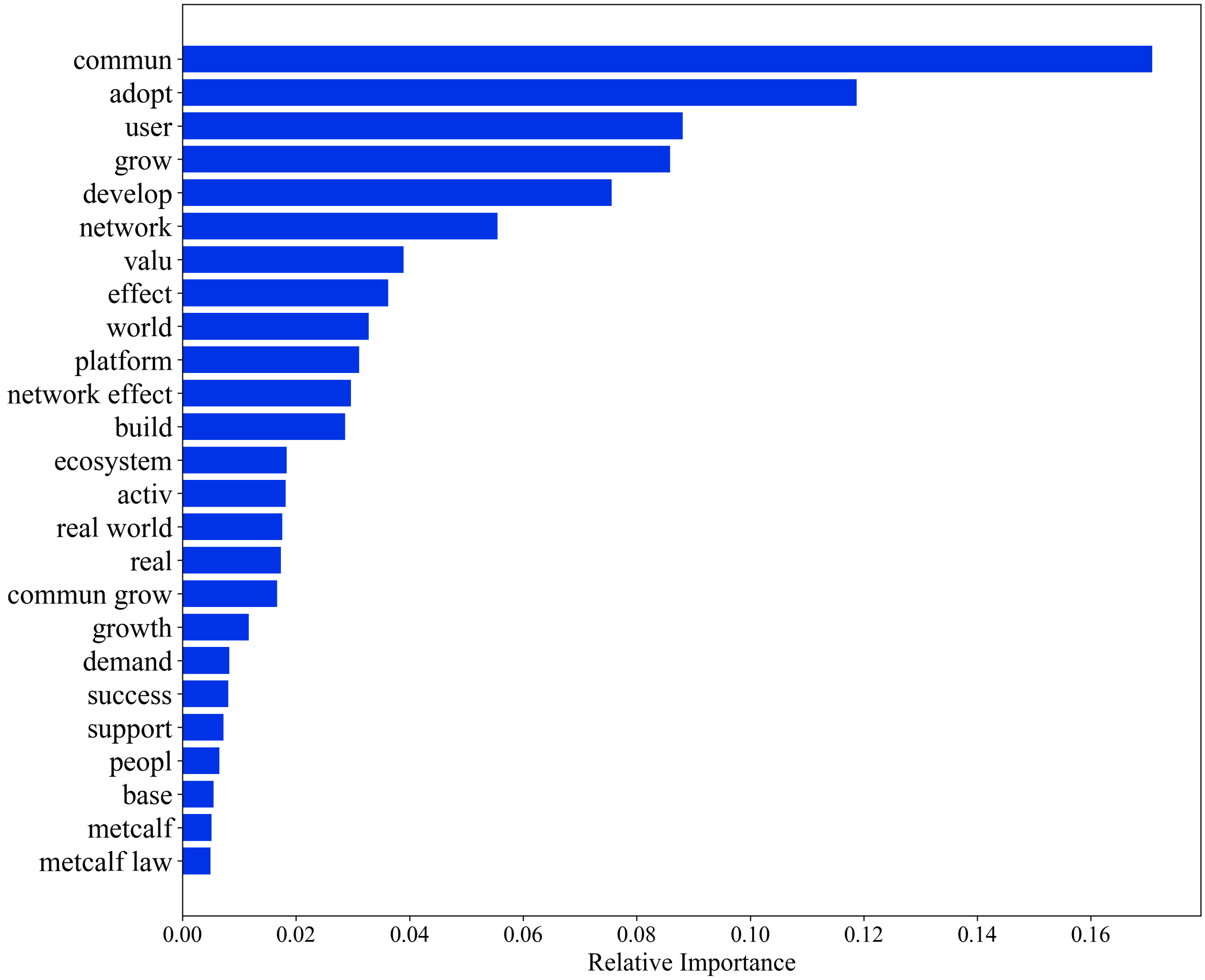


Figure 9. Random Forest Feature Importance for Classifying Reddit Comments Based on the Network Effect.

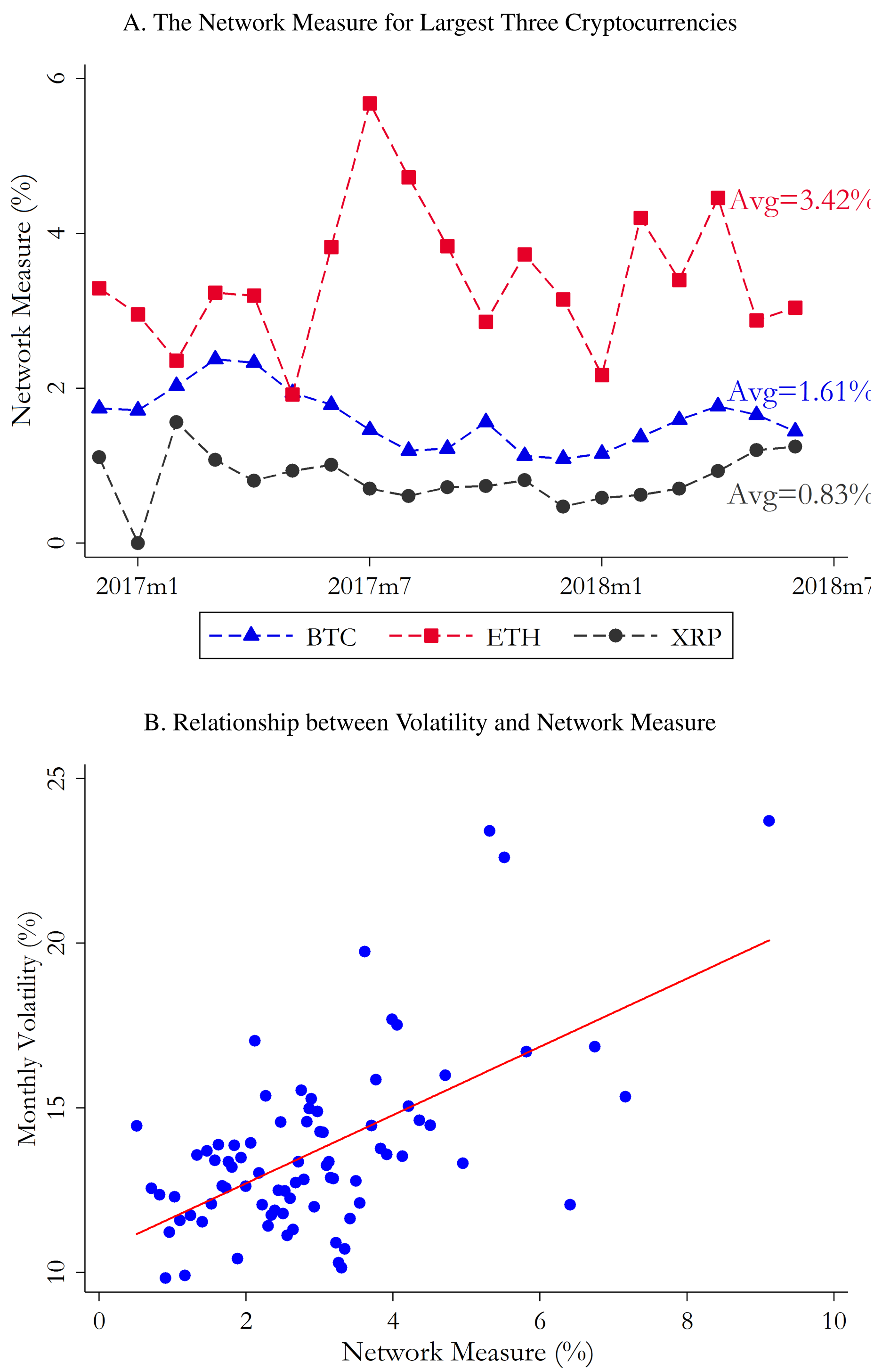


Figure 10. Percentage of the Network-Based Comments in Reddit Pages of Cryptocurrencies. Panel A shows the percentage of comments using the word *community* in the Reddit pages of Bitcoin (*/r/Bitcoin* and */r/btc*), Ethereum (*/r/ethereum*), and Ripple (*/r/ripple*). Panel B shows the binscatter of average monthly volatility of cryptocurrencies on average monthly percentage of comments using the word *community*. Volatility is calculated using daily returns. Currency-month observations with less than 100 comments are excluded.

Table I. Number of Cryptocurrencies in the Sample Over Time. This table summarizes the number of cryptocurrencies as well as the equal-weighted and volume-weighted daily returns of the market portfolio for cryptocurrencies in the sample. The sample includes all cryptocurrencies in *CoinAPI* or *Kaiko* data conditional on having an average daily transaction of at least 1,000 trades in the prior week and average quoted price of at least ten Satoshis in the prior week. Cryptocurrencies defined as derivatives of other currencies such as Bitcoin futures are excluded. Pegged digital currencies such as Tether are excluded. The sample period is from January 1, 2017 to June 30, 2018.

Month	N Currencies	EW Daily Ret (%)	VW Daily Ret (%)
2017M01	50	0.60	0.03
2017M02	59	0.21	0.86
2017M03	237	3.52	2.14
2017M04	132	1.74	2.21
2017M05	133	2.96	3.99
2017M06	193	1.45	1.52
2017M07	218	-0.86	0.40
2017M08	237	1.84	3.04
2017M09	275	-0.03	0.20
2017M10	264	-0.37	1.37
2017M11	315	1.34	3.71
2017M12	372	3.92	4.70
2018M01	549	-0.06	0.50
2018M02	499	-0.92	0.41
2018M03	512	-1.87	-1.22
2018M04	599	2.17	2.57
2018M05	574	-0.92	-0.38
2018M06	554	-1.58	-0.67

Table II. Commonalities in Order Imbalance of Cryptocurrencies Listed on the Same Exchange. This table reports the estimates for a panel regression of the form:

$$OIB_{i,k,t} = \beta_0 + \beta_1 \overline{OIB}^{Cur}_{i,k,t} + \beta_2 \overline{OIB}^{Exch}_{i,k,t} + \beta_3 \overline{OIB}^{Mkt}_{i,k,t} + \delta_{i,k} + \varepsilon_{i,k,t}$$

where $OIBi_{i,k,t}$ denotes the order imbalance for currency i on exchange k on day t, $\overline{OIB}^{Cur}_{i,k,t}$ is the average order imbalance of currency i on all other exchanges excluding k, $\overline{OIB}^{Exch}_{i,k,t}$ is the average imbalance of other currencies on exchange k excluding currency i, $\overline{OIB}^{Mkt}_{i,k,t}$ is the average imbalance of the rest of the market, and $\delta_{i,k}$ is currency-exchange fixed effect. Panel A reports the results for the order imbalance based on the number of transactions ($OIBNUM$) and Panel B based on the dollar volume ($OIBVOL$) as detailed in Section III. The sample period is from January 1, 2017 to June 30, 2018. Standard errors are two-way clustered at currency and exchange level. t-statistics are reported in parentheses. *p<.05, **p<.01, ***p<.001.

A. Order Imbalance Based on the Number of Transactions

	(1)	(2)	(3)	(4)	(5)
OIBNUM^{Cur}	0.125**			0.0925*	0.102*
	(2.96)			(2.45)	(2.35)
OIBNUM^{Exch}		0.309***		0.299***	0.298***
		(7.11)		(5.85)	(5.72)
OIBNUM^{Mkt}			0.0921*	-0.00342	-0.00424
			(2.24)	(-0.16)	(-0.19)
L.OIBNUM^{Cur}					-0.00791
					(-0.77)
F.OIBNUM^{Cur}					-0.0126
					(-1.83)
Curr-Exch FE	Yes	Yes	Yes	Yes	Yes
Observations	111528	111528	111528	111528	107899
R^2	0.021	0.102	0.013	0.110	0.111
Curr & Exch Cluster	Yes	Yes	Yes	Yes	Yes

B. Order Imbalance Based on the Dollar Volume

	(1)	(2)	(3)	(4)	(5)
$\text{OIBVOL}^{\text{Cur}}$	0.219***			0.175***	0.192***
	(5.00)			(4.48)	(4.79)
$\text{OIBVOL}^{\text{Exch}}$		0.271***		0.235***	0.233***
		(9.68)		(6.59)	(6.37)
$\text{OIBVOL}^{\text{Mkt}}$			0.175***	0.0115	0.0112
			(5.70)	(0.63)	(0.60)
$\text{L.OIBVOL}^{\text{Cur}}$					-0.0235**
					(-3.54)
$\text{F.OIBVOL}^{\text{Cur}}$					-0.0226**
					(-3.62)
Curr-Exch FE	Yes	Yes	Yes	Yes	Yes
Observations	111516	111516	111516	111516	107879
R^2	0.053	0.078	0.033	0.109	0.112
Curr & Exch Cluster	Yes	Yes	Yes	Yes	Yes

Table III. Connectivity and Comovement in Returns and Order Imbalance. Panel A reports estimates of a panel regression where the dependent variable is the within-month standardized pairwise correlation of the market model return residuals:

$$Corr_{i,j,t} = \beta_0 + \beta_1 Connectivity_{i,j,t-1} + \beta^{Char} Similarity^{Char}_{i,j,t-1} + \delta_t + \varepsilon_{i,j,t}$$

The market return is the volume-weighted returns of all cryptocurrencies, weighted by currencies' previous 24-hour volume. Market model residuals are estimated using a recursive time-series regression of individual currency returns on the market return. $Corr_{i,j,t}$ measures the within-month correlation of the return residuals for all pairwise combinations. Market betas, return residuals, and pairwise correlations are calculated using rolling windows of 24-hour returns that move forward every hour. The $Connectivity$ measure is constructed as described in Section II.B.1. $Similarity^{Char}$ is a vector of binary and continuous variables which directly control for similarity in characteristics as detailed in Section II.B.2. Return correlations, the $Connectivity$ measure, and the continuous similarity variables are standardized by subtracting the mean and dividing by the standard deviation. Panel B reports similar results for pairwise correlations in order imbalance. Standard errors are clustered at the dyadic level. The sample period is from January 1, 2017 to June 30, 2018. t-statistics are reported in parentheses. *p<.05, **p<.01, ***p<.001.

A. Comovement in Returns

	All Currencies			Large Currencies		
	(1)	(2)	(3)	(4)	(5)	(6)
Connectivity	0.189***	0.174***	0.172***	0.213***	0.190***	0.187***
	(20.21)	(19.72)	(19.82)	(11.36)	(12.05)	(12.22)
$\text{Similarity}^{\text{Volume}}$		0.068***	0.068***		0.204***	0.199***
		(7.67)	(7.58)		(4.00)	(3.92)
$\text{Similarity}^{\text{NExch}}$		0.040*	0.039*		0.057*	0.057*
		(2.30)	(2.29)		(2.51)	(2.54)
$\text{Similarity}^{\text{CoinToken}}$			0.037**			0.042
			(2.99)			(1.72)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	447771	444592	444592	62549	61537	61537
Adjusted R^2	.179	.185	.185	.176	.186	.187
Dyadic Clustering	Yes	Yes	Yes	Yes	Yes	Yes

B. Comovement in Order Imbalance

	All Currencies			Large Currencies		
	(1)	(2)	(3)	(4)	(5)	(6)
Connectivity	0.130***	0.122***	0.121***	0.129***	0.130***	0.128***
	(16.29)	(15.34)	(15.37)	(8.02)	(8.30)	(8.23)
$\text{Similarity}^{\text{Volume}}$		0.035***	0.034***		0.147***	0.142***
		(5.33)	(5.29)		(3.92)	(3.75)
$\text{Similarity}^{\text{NExch}}$		-0.039*	-0.040*		-0.051*	-0.051*
		(-2.43)	(-2.43)		(-2.13)	(-2.12)
$\text{Similarity}^{\text{CoinToken}}$			0.017			0.043
			(1.23)			(1.60)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	340505	334410	334410	55097	53202	53202
Adjusted R^2	.06	.064	.062	.049	.055	.051
Dyadic Clustering	Yes	Yes	Yes	Yes	Yes	Yes

Table IV. Connectivity, Technological Characteristics, and Return Comovement. This table reports a similar regression as in Table III separately for the subset of coins and tokens and controls for coin- and token-specific characteristics:

$$Corr_{i,j,t} = \beta_0 + \beta_1 Connectivity_{i,j,t-1} + \beta^{Char} Similarity^{Char}_{i,j,t-1} + \delta_t + \varepsilon_{i,j,t}$$

The dummy variable $Similarity^{ProofType}$ takes the value of one if the pair has the same consensus mechanisms, $Similarity^{HashAlgo}$ takes the value of one if the pair has the same hashing algorithm, and $Similarity^{Fork}$ takes the value of one if the two coins are forks on the same blockchain. $Similarity^{Platform}$, $Similarity^{Industry}$, and $Similarity^{TokenType}$ are dummy variables that take the value of one if the token pair are developed on top of the same blockchain, are in the same industry, and are both utility or equity tokens respectively. Panel B reports estimates of a similar regression for 37 large coins with available data on a larger set of characteristics. These characteristics include end-of-the-month block time and mining difficulty as well as the blockchain maturity, which is measured as the passage of time from the genesis block. Standard errors are clustered at the dyadic level. The sample period is from January 1, 2017 to June 30, 2018. t-statistics are reported in parentheses. *p<.05, **p<.01, ***p<.001.

A. Coin and Token Characteristics

	Coin Pairs			Token Pairs		
	(1)	(2)	(3)	(4)	(5)	(6)
Connectivity	0.196***		0.190***	0.194***		0.180***
	(11.99)		(11.32)	(16.97)		(16.21)
$\text{Similarity}^{\text{ProofType}}$		0.079*	0.063*			
		(2.29)	(2.06)			
$\text{Similarity}^{\text{HashAlgo}}$		0.055	0.028			
		(1.73)	(0.98)			
$\text{Similarity}^{\text{Fork}}$		0.081	0.086			
		(1.21)	(1.43)			
$\text{Similarity}^{\text{Volume}}$			0.035*			0.076***
			(2.38)			(6.91)
$\text{Similarity}^{\text{Platform}}$					0.032	0.004
					(1.07)	(0.17)
$\text{Similarity}^{\text{Industry}}$					0.121**	0.098**
					(3.07)	(2.97)
$\text{Similarity}^{\text{TokenType}}$					-0.060	-0.044
					(-1.89)	(-1.63)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	82588	82588	82588	111386	111386	111386
Adjusted R^2	0.180	0.143	0.182	0.206	0.168	0.210
Double Cluster	Yes	Yes	Yes	Yes	Yes	Yes

B. More Characteristics for Major Coins

	(1)	(2)	(3)
Connectivity	0.189***	0.171***	0.167***
	(6.77)	(5.32)	(5.26)
$Similarity^{Volume}$		0.036	0.043
		(1.09)	(1.27)
$Similarity^{NExch}$		0.140*	0.119*
		(2.32)	(2.05)
$Similarity^{ProofType}$		-0.052	-0.059
		(-0.94)	(-1.23)
$Similarity^{HashAlgo}$		-0.034	-0.069
		(-0.55)	(-1.04)
$Similarity^{Fork}$		-0.022	-0.005
		(-0.25)	(-0.06)
$Similarity^{BlockTime}$			0.034
			(1.46)
$Similarity^{Maturity}$			0.029
			(1.19)
$Similarity^{Difficulty}$			0.053
			(1.77)
Time FE	Yes	Yes	Yes
Observations	5,687	5,687	5,687
Adjusted R^2	0.13	0.153	0.158
Dyadic Clustering	Yes	Yes	Yes

Table V. Explanatory Power of Connectivity and Other Characteristics for Return Correlations. Panel A reports the average realized pairwise correlations for ten deciles sorted on the realized correlations ($Realized$), fitted values of the regression in Equation (3) with one-month lagged connectivity as the explanatory variable ($Connectivity$), and the fitted values using all other characteristics excluding connectivity ($Characteristics$). At each point in time, cryptocurrency pairs are sorted into deciles of the fitted values, and the average realized correlations are calculated for each decile. The values for each decile are then averaged over time. The last row reports the spread between deciles one and ten obtained from the fitted values as the percentage of the actual spread. Panel B reports the contribution of different variables to the cross-sectional adjusted R^2. The left side of panel B reports the average R^2s for monthly cross-sectional regressions of return correlations on lagged connectivity or other characteristics. The right side of panel B shows the average incremental contribution of each set of variables when all other variables are already included in the regression. The sample period is from January 1, 2017 to June 30, 2018.

A. Decile Spread

Decile	Realized	Connectivity	Characteristics
1	-1.65	-0.11	-0.05
2	-0.90	-0.01	0.05
3	-0.54	-0.03	0.09
4	-0.26	0.05	0.11
5	-0.00	0.08	0.14
6	0.25	0.19	0.11
7	0.51	0.33	0.16
8	0.80	0.44	0.16
9	1.17	0.54	0.23
10	1.86	0.51	0.25
Percentage of Actual Spread		**17.81%**	**8.55%**

B. Contribution to Adjusted R^2

	R^2				Incremental R^2			
	All	Large	Coins	Tokens	All	Large	Coins	Tokens
Connectivity	5.7%	5.5%	5.9%	4.4%	3.9%	3.7%	3.7%	3.6%
Size	1.5%	2.0%	1.3%	1.4%	0.5%	1.2%	0.4%	0.4%
Technology	1.1%	1.7%	0.6%	0.9%	1.0%	1.7%	0.4%	0.7%
Volume	0.9%	0.5%	0.7%	1.6%	0.2%	0.2%	0.2%	0.6%

Table VI. Cross-Sectional Fama-MacBeth Regressions of Cryptocurrency Returns on Connected Portfolio Returns. This table shows the results of cross-sectional Fama-MacBeth regressions of cryptocurrencies returns on contemporaneous returns of their connected portfolio. The returns of the connected portfolio are calculated as described in Section IV.D. The results are estimated using overlapping returns with rolling windows that move forward every hour. The reported t-statistics in parentheses are based on autocorrelation-consistent Newey-West standard errors. For example, the standard errors for the daily (weekly) results are adjusted using the Newey-West procedure with up to 24 (168) lags. Panel A shows the results for the sample period from January 1, 2017 to June 30, 2018 and panel B for the first half of 2018. *p<.05, **p<.01, ***p<.001.

	A. All Sample							
	12hr	1d	2d	3d	4d	5d	6d	7d
Connected Ret	0.30***	0.32***	0.38***	0.41***	0.39**	0.40	0.53**	0.57**
	(7.60)	(5.31)	(3.89)	(3.98)	(2.72)	(1.87)	(2.79)	(2.84)
Constant	0.0026*	0.0044	0.010	0.013	0.017	0.025	0.030	0.043*
	(2.14)	(1.70)	(1.80)	(1.58)	(1.55)	(1.73)	(1.83)	(2.17)
Observations	2623410	2578365	2506383	2443071	2386977	2335407	2287855	2244988

	B. 2018							
	12hr	1d	2d	3d	4d	5d	6d	7d
Connected Ret	0.50***	0.56***	0.64***	0.70***	0.77***	0.79***	0.84***	0.87***
	(17.59)	(13.22)	(10.78)	(9.90)	(9.94)	(9.64)	(10.60)	(10.52)
Constant	-0.0038*	-0.0060	-0.015*	-0.025**	-0.023*	-0.022	-0.021	-0.023
	(-2.28)	(-1.85)	(-2.50)	(-2.72)	(-2.01)	(-1.60)	(-1.36)	(-1.39)
Observations	1539861	1518133	1482373	1451312	1424147	1399895	1377322	1357300

Table VII. Daily Returns of Zero-Cost Trading Strategy Based on the Connected Portfolio Returns. This table shows the average daily returns of trading strategies that are long currencies in the highest decile and short in the lowest decile of previous J days connected portfolio returns and hold the portfolio for the next K days. The portfolios are rebalanced every hour, and at each point in time, the strategy holds $24K$ overlapping portfolios. The returns of the connected portfolio are calculated as described in Section IV.D. The sample period is from January 1, 2017 to June 30, 2018. The standard errors are adjusted using the Newey-West method based on the holding horizon. For example, the standard errors for the daily (weekly) results are adjusted using the Newey-West procedure with up to 24 (168) lags. *p<.05, **p<.01, ***p<.001.

J	*K=*	*12hr*	*1d*	*2d*	*3d*	*4d*	*5d*	*6d*	*7d*
12hr		0.0064	0.0052	0.0048	0.0029	0.0024	0.0027	0.0031	0.0028
		(3.53)	(3.36)	(3.92)	(2.92)	(2.74)	(3.06)	(3.58)	(3.16)
1d		0.0065	0.0071	0.0052	0.0036	0.0027	0.0030	0.0037	0.0031
		(3.31)	(3.68)	(3.25)	(2.67)	(2.34)	(2.59)	(3.30)	(2.87)
2d		0.0104	0.0075	0.0040	0.0033	0.0038	0.0045	0.0044	0.0034
		(4.80)	(3.58)	(1.97)	(1.82)	(2.30)	(2.88)	(3.20)	(2.46)
3d		0.0068	0.0046	0.0028	0.0029	0.0044	0.0045	0.0038	0.0026
		(3.16)	(2.11)	(1.23)	(1.40)	(2.30)	(2.49)	(2.31)	(1.59)
4d		0.0056	0.0049	0.0055	0.0062	0.0065	0.0059	0.0049	0.0040
		(2.55)	(2.19)	(2.36)	(2.69)	(2.93)	(2.76)	(2.43)	(1.97)
5d		0.0073	0.0072	0.0073	0.0073	0.0067	0.0056	0.0049	0.0038
		(3.03)	(2.77)	(2.71)	(2.80)	(2.73)	(2.45)	(2.20)	(1.71)
6d		0.0082	0.0080	0.0074	0.0062	0.0052	0.0047	0.0037	0.0029
		(3.22)	(2.90)	(2.57)	(2.29)	(2.02)	(1.88)	(1.59)	(1.32)
7d		0.0077	0.0074	0.0061	0.0044	0.0041	0.0036	0.0029	0.0025
		(3.08)	(2.68)	(2.18)	(1.69)	(1.60)	(1.45)	(1.25)	(1.16)

Table VIII. Exchange Listing and Changes in Comovement. This table presents estimates of the regression of within-month correlation of market model residuals on a treatment dummy variable, $Treated_{i,j}$, which takes the value of one if currency i is listed on an exchange that already lists currency j or vice versa. For each treated pair, a control group is matched by finding the ten closest currency pairs using the Mahalanobis distance in the pairwise connectivity and the monthly trading volume of the two currencies the month prior to listing. This table reports estimates of a regression of the form:

$$\begin{aligned} Corr_{i,j,t} =& \beta_0 + \beta_1 Treated_{i,j} + \sum_{t=-1}^{3} \gamma_t Treated_{i,j} * M_t + \beta^{Char} Similarity^{Char}_{i,j,t-1} \\ &+ \sum_{t=-1}^{3} \gamma^{Char}_t Similarity^{Char}_{i,j,t-1} * M_t + \delta_{c,t} + \delta_{i,j} + \varepsilon_{i,j,t} \end{aligned}$$

where $\delta_{c,t}$ is the cohort-time fixed effect and $\delta_{i,j}$ is the currency pair fixed effect. Standard errors are clustered at the dyadic level. t-statistics are reported in parentheses. *p<.05, **p<.01, ***p<.001.

	(1)	(2)	(3)	(4)
Treated	-0.058	-0.049	-0.050	-0.049
	(-1.35)	(-1.13)	(-1.15)	(-1.14)
Treated*M=-1	0.026	0.017	0.019	0.018
	(0.58)	(0.37)	(0.42)	(0.39)
Treated*M=0	0.008	-0.000	0.007	0.010
	(0.17)	(-0.01)	(0.16)	(0.22)
Treated*M=1	0.142*	0.135*	0.129*	0.126*
	(2.46)	(2.28)	(2.18)	(2.12)
Treated*M=2	0.192*	0.185*	0.184*	0.183*
	(2.54)	(2.42)	(2.36)	(2.35)
Treated*M=3	0.301***	0.294***	0.294***	0.294***
	(3.42)	(3.37)	(3.33)	(3.33)
$\text{Similarity}^{\text{SameCoinToken}}$		0.031***	0.031***	0.031***
		(3.73)	(3.75)	(3.76)
$\text{Similarity}^{\text{Volume}}$		-0.008	-0.021	-0.020
		(-0.56)	(-1.39)	(-1.33)
$\text{Similarity}^{\text{SameNExch}}$		-0.001	-0.001	-0.003
		(-0.02)	(-0.02)	(-0.11)
Cohort-Time FE	Yes	Yes	Yes	Yes
Pair FE	Yes	Yes	Yes	Yes
Observations	225011	224466	224466	224466
Dyadic Clustering	Yes	Yes	Yes	Yes
$\text{Similarity}^{\text{Volume}}*\text{M}$	No	No	Yes	Yes
$\text{Similarity}^{\text{SameNExch}}*\text{M}$	No	No	No	Yes

Table IX. 2SLS Estimation of the Effect of Shutdown of Chinese Exchanges on Return Comovement. This table presents two-stage least-squares regression estimates for the effect of changes in the connectivity caused by shutdown of Chinese exchanges on changes in return comovement in a regression of the form:

$$\Delta Corr_{i,j} = \beta_0 + \beta_1 \Delta \widehat{Connectivity}_{i,j} + \beta^{Char} \Delta Similarity^{Char}_{i,j} + \delta_c + \varepsilon_{i,j}$$

where $\Delta Corr_{i,j}$ is change in the average monthly return correlations from the three months prior to the shutdown to three months after, δ_c is the matching cohort fixed effect, and $\Delta \widehat{Connectivity}_{i,j}$ represents the fitted values from the first stage regression of the form:

$$\Delta Connectivity_{i,j} = \gamma_0 + \gamma_1 Connectivity^{Chinese}_{i,j} + \gamma^{Char} \Delta Similarity^{Char}_{i,j} + \delta_c + \varepsilon_{i,j}$$

The variable $Connectivity^{Chinese}$ captures the contribution of Chinese exchanges to the connectivity measure and is calculated as described in Section V.B. For each treated pair, a control group is matched by finding the ten closest currency pairs based on the Mahalanobis distance in the pairwise connectivity and the monthly trading volume of both currencies the month before the shutdown. Standard errors are clustered at the dyadic level. t-statistics are reported in parentheses. *p<.05, **p<.01, ***p<.001.

	(1)	(2)	(3)
	ΔCorr	ΔCorr	ΔCorr
ΔConnectivity	0.188***	0.174**	0.173**
	(3.34)	(2.78)	(2.79)
ΔSimilarity$^{\text{Volume}}$		0.032	0.034
		(1.09)	(1.11)
ΔSimilarity$^{\text{NExch}}$			-0.003
			(-0.18)
Observations	40887	40887	40887
First-Stage F statistics	21.42	21.22	21.09

Table X. Persistence and Consistency of the Community Index. Panel A presents the monthly autocorrelation of the percentage of currency-specific comments that use the word *community* on the Reddit pages of different cryptocurrencies. The analysis is at the currency-month level and the standard errors are clustered at the currency level. Panel B presents the average monthly percentage usage of the word *community* for the ten largest cryptocurrencies based on the market cap as of June 30, 2018 as well as an indicator for whether these currencies are platform tokens. A cryptocurrency is designated as high-community-based if the average monthly usage of the word *community* is above the median of the sample. Currency-month observations with less than 100 comments are excluded. t-statistics are reported in parentheses. *p<.05, **p<.01, ***p<.001.

A. Persistence of the Community Measure

	(1)	(2)	(3)
L.Perc_COMM	0.631***	0.535***	0.505***
	(19.24)	(12.73)	(8.73)
L2.Perc_COMM		0.211***	0.150***
		(5.92)	(3.45)
L3.Perc_COMM			0.157***
			(3.78)
Constant	0.009***	0.007***	0.005***
	(12.74)	(8.19)	(4.82)
Time FE	Yes	Yes	Yes
Observations	2759	2253	1871
Adjusted R^2	0.413	0.461	0.488

B. The Community Measure and the Platform Token Indicator

N	Ticker	Perc_Comm	Hi_Comm	PlatToken	
1	ETH	3.4	1	1	✓
2	ADA	2.5	1	1	✓
3	IOTA	2.3	1	0	✗
4	EOS	2.2	1	1	✓
5	BTC	1.6	0	0	✓
6	BCH	1.6	0	0	✓
7	TRX	1.5	0	0	✓
8	XLM	1.5	0	0	✓
9	LTC	1.3	0	0	✓
10	XRP	0.8	0	0	✓

Table XI. Amplification of the Demand Shocks for High-Community-Based Cryptocurrencies. This table estimates a panel regression of the form:

$$\begin{aligned} Corr_{i,j,t} =& \beta_0 + \beta_1 Connectivity_{i,j,t-1} + \beta_2 HI\_COMM_{i,j} + \beta_3 Connectivity_{i,j,t-1} * HI\_COMM_{i,j} \\ &+ Controls + \delta_t + \varepsilon_{i,j,t} \end{aligned}$$

where $HI\_COMM_{i,j}$ is a dummy variable that takes the value of one if the average monthly percentage of the Reddit comments that use of the word *community* is above the median for both currencies i and j and zero if below the median. Currency-month observations with less than 100 comments are excluded. Standard errors are clustered at the dyadic level. t-statistics are reported in parentheses. *p<.05, **p<.01, ***p<.001.

A. Based on Machine Learning Classification

	(1)	(2)	(3)	(4)
Connectivity	0.151***	0.137***	0.138***	0.139***
	(10.75)	(9.53)	(7.87)	(7.99)
HI_COMM	0.184***	0.176***	0.177***	0.157**
	(3.94)	(3.78)	(3.79)	(3.16)
Connectivity*HI_COMM	0.066**	0.070***	0.070***	0.075**
	(3.29)	(3.47)	(3.49)	(3.17)
$\text{Similarity}^{\text{Volume}}$		0.074***	0.074***	0.075***
		(4.58)	(4.61)	(3.73)
$\text{Similarity}^{\text{CoinToken}}$		0.022		
		(1.83)		
$\text{Connectivity*Similarity}^{\text{Volume}}$			0.000	
			(0.03)	
Time FE	Yes	Yes	Yes	Yes
Observations	80940	80940	80940	40123
Adjusted R^2	0.229	0.230	0.230	0.230
Dyadic Clustering	Yes	Yes	Yes	Yes

B. Based on Frequency of the Word "Community"

	(1)	(2)	(3)	(4)
Connectivity	0.153***	0.140***	0.136***	0.139***
	(9.16)	(8.29)	(6.74)	(8.15)
HI_COMM	0.082	0.088	0.087	0.025
	(1.60)	(1.75)	(1.73)	(0.53)
Connectivity*HI_COMM	0.057**	0.057**	0.059**	0.073**
	(2.62)	(2.62)	(2.72)	(2.89)
$\text{Similarity}^{\text{Volume}}$		0.088***	0.088***	0.084***
		(5.61)	(5.66)	(4.55)
$\text{Similarity}^{\text{CoinToken}}$		0.023		
		(1.39)		
$\text{Connectivity*Similarity}^{\text{Volume}}$			0.005	
			(0.31)	
Time FE	Yes	Yes	Yes	Yes
Observations	83315	83315	83315	44086
Adjusted R^2	0.223	0.225	0.225	0.230
Dyadic Clustering	Yes	Yes	Yes	Yes

Internet Appendix

IA.A. Cleaning and Filtering the Trading and Price Data

To ensure data quality, *CoinAPI*, *Kaiko*, and *CoinMarketCap* datasets are merged using currencies' tickers and exhaustively cross-checked for any discrepancies. Whenever discrepancies of more than 10% are observed in daily price or trading volume of a given cryptocurrency across these datasets, I investigate the causes and either fix the issue or drop the currency.

First, currencies are identified by a ticker on each exchange. However, these tickers are not necessarily unique across exchanges, especially for currencies with similar names. For example, the correct ticker for the currency Dash is DASH, but it is listed as DSH on certain exchanges, a ticker which belongs to a different currency named Dashcoin. These discrepancies in the reported tickers are fixed. Second, for the majority of exchanges, the reported volume specifies the trading volume of the base currency (the first currency in a currency pair quotation). However, for a few exchanges, the reported volume specifies the trading volume of the quote currency without indicating this difference. These observations are identified based on discrepancies in the aggregate volume and then fixed. Third, there are some data inputs for exchanges that went out of business after they ceased operation. Such observations are removed. Finally, Makarov and Schoar (2018) find that buy trades are reported as sell trades and vice versa for Bithumb and Quoine exchanges and that Bithumb prices are reported in local Korean time. I correct for these issues. In a similar fashion, I find that Yobit buy versus sell trade indicators must be corrected in the same way as Bithumb and Quoine.

To ensure only liquid currencies with frequently updated prices are included in my sample, at each point in time, I filter out currency pairs that have average daily transactions of less than 1,000 trades in the prior week. The results are robust to different cutoffs for excluding illiquid currencies, such as average daily transactions of 500 or 2,000. I also excluded currencies defined as derivatives of other currencies, such as Bitcoin futures on Bitmex. Moreover, I use pegged digital currencies such as Tether to price other cryptocurrencies, but I excluded them from price analysis because their pricing is inherently different due to their pegged nature. Finally, prices quoted in Bitcoin often have a minimum tick level of a "Satoshi" (10^{-8} BTC). Even prices quoted in other cryptocurrencies such as Ether follow the same rule. I require the average quoted price in the prior week to be at least ten ticks, or 10^{-7}.

Table IAI. Technological Features of Cryptocurrencies Used in This Paper.

Characteristics	Description	Variable Type
Coins vs. Tokens	Coins are closer to traditional currencies and mainly function as a medium of exchange. Tokens can also be used as a medium of exchange, but they can offer other functionalities such as representing a tradable asset or utility. Tokens provide the right to future use of a specific product or service developed by the issuer. Data source: *CoinMarketCap*	Dummy
Hashing Algorithm	The cryptographic algorithm used by different coins. For example, Bitcoin uses a hashing algorithm labeled "SHA-256," whereas Litecoin, another large digital coin, uses the hashing algorithm called "Scrypt." SHA-256 algorithm is more complex and is believed to provide a higher degree of data security, while Scrypt algorithm is faster and is more energy efficient. This makes Litecoin a better choice for day-to-day transactions due to a faster verification process. The main hashing algorithms in the sample are SHA-256, Scrypt, X11, Cryptonight, and Ethash. Data source: *Blockchainio.io*, *Allcoinsnews.com*, and *Cryptopia.co.nz*.	Categorical
Consensus Mechanism	Is the mechanism through which the network reaches a consensus about the information being added to the ledger. Such information could validate transactions on the blockchain, ensuring that the next block being added represents the most current transactions on the network. For example, Bitcoin uses Proof-of-Work (POW) while NEO, another large digital coin, uses Proof-of-Stake (POS). The POW consensus mechanism requires continually increasing computing power and electricity, but the POS algorithm requires significantly less resources. Data source: *Blockchainio.io*, *Allcoinsnews.com*, and *Cryptopia.co.nz*.	Categorical
Forks on the Same Blockchain	Forks are variations of the same currency, like a spin-off, with modifications to the original protocol through adding new features. For example, Bitcoin Cash and Bitcoin Gold are two of the many forks on the Bitcoin blockchain, while Ethereum Classis is a fork on the Ethereum blockchain. Data source: *Mapofcoins.com* and *Blockchainalgo.com*	Categorical

Block Time	Represents the average time it takes for the network to create a new block on the blockchain. Differences in the block time can significantly affect the transaction speed on a blockchain. For example, the block time for bitcoin is 10 around minutes as of June 2018, but the while for Ethereum it is 14 and 15 seconds. Data source: *BitInfoCharts.com*	Continuous
Time from the Genesis Block	Time from the genesis block can represent the maturity of a blockchain. For example, the genesis block on the Bitcoin blockchain was created in January 2009, while the genesis block of Ethereum was incepted in July 2015. Data source: *BitInfoCharts.com*	Continuous
Hash Rate	Shows the number of hashes per second the network of a blockchain is performing. The hash rate reflects the difficulty of mining new coins. As the difficulty increase, the electricity and computing power costs of mining new currencies increase. Data source: *BitInfoCharts.com*	Continuous
Token Industry	Tokens can represent projects that are classified into different industries. Top industries in the sample are Blockchain Infrastructure, Financial Services, Exchanges and Wallets, and Computing and Data Storage. Data source: *Icorating.com*	Categorical
Token Platform	Tokens differ in their underlying platforms. While the majority of tokens are smart contracts based on the Ethereum blockchain, some tokens are based on other platforms such as NEO or Waves. Data source: *CoinMarketCap*	Categorical
Equity vs. Utility Token	Tokens are divided into equity tokens, representing ownership of an asset, or utility tokens, providing future access to a product or service. These two types have different legal requirements and could be susceptible to different types of legal shocks. Data source: *Icorating.com*	Categorical

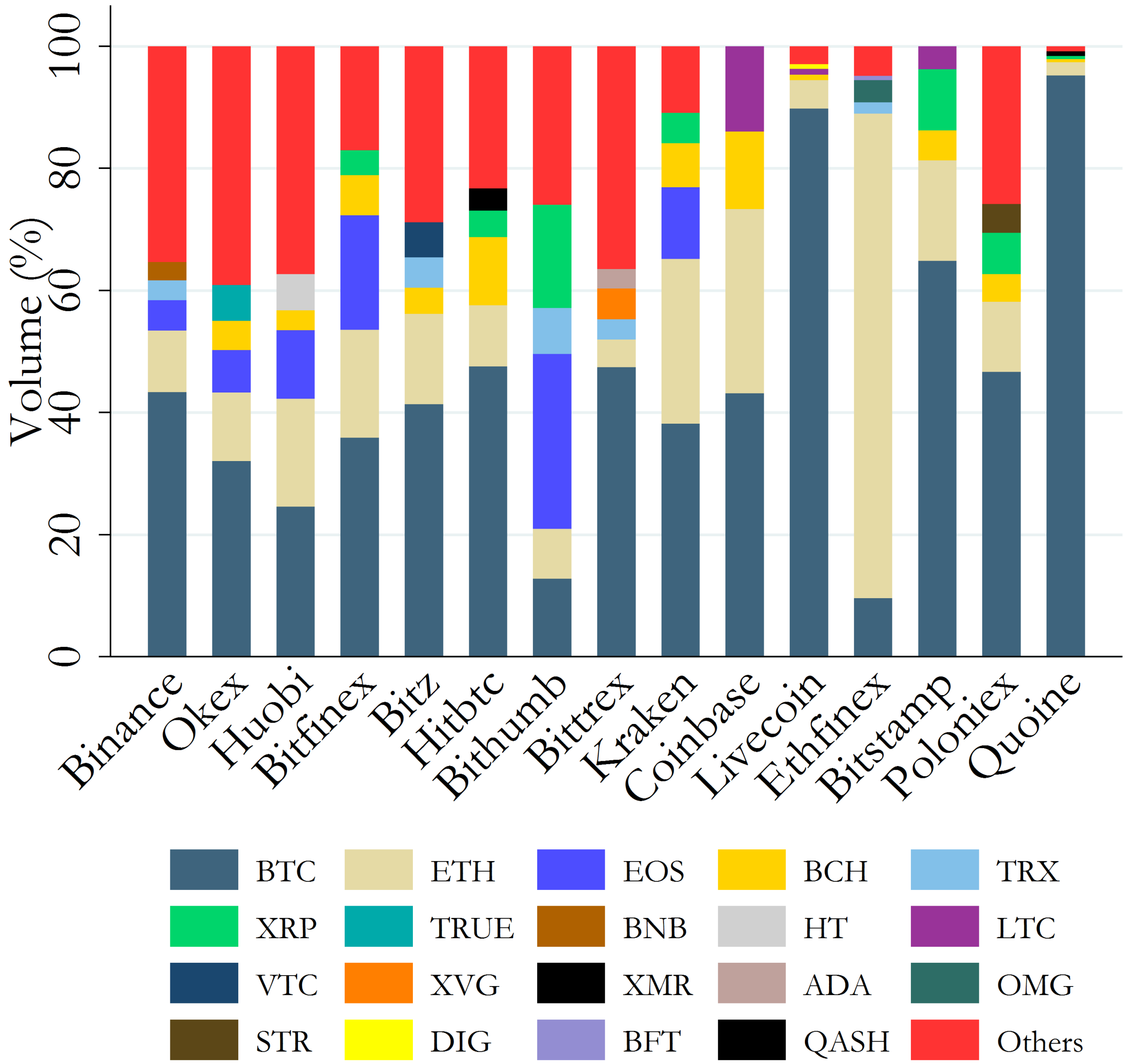
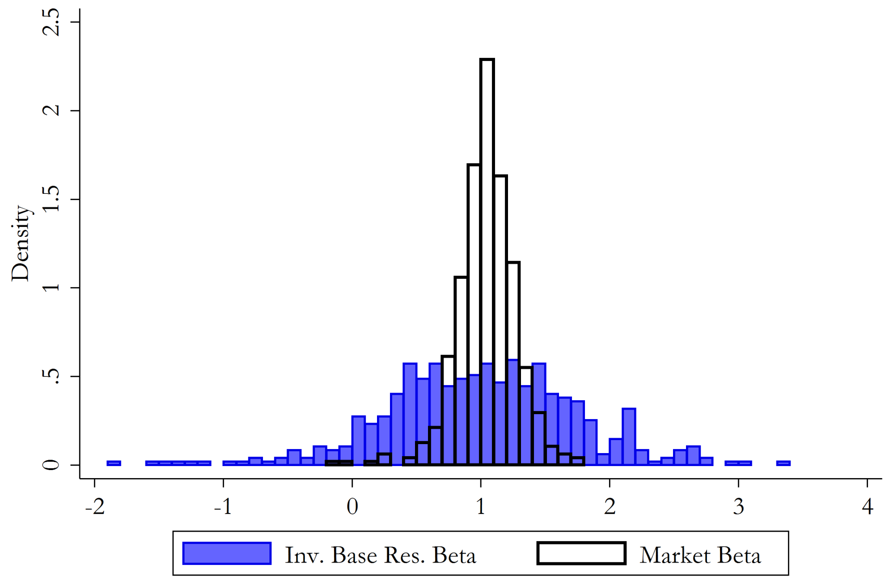


Figure IA1. Variation in Percentage of Trading Volume of Major Cryptocurrencies on Different Exchanges. This figure shows the variation in trading volume of major cryptocurrencies across different exchanges available in *CoinAPI* and *Kaiko* data in the second quarter of 2018. Major exchanges are selected as the top 15 exchanges in terms of dollar trading volume in this period. The figure plots the trading volume of top currencies on each exchange as the percentage of total volume of that exchange. The exchanges are sorted based on trading volume, where the volume decreases from left to right. Pegged cryptocurrencies are excluded.

A. Investor Base Returns are Orthogonalized Relative to Market Returns



B. Market Returns are Orthogonalized Relative to Investor Base Returns

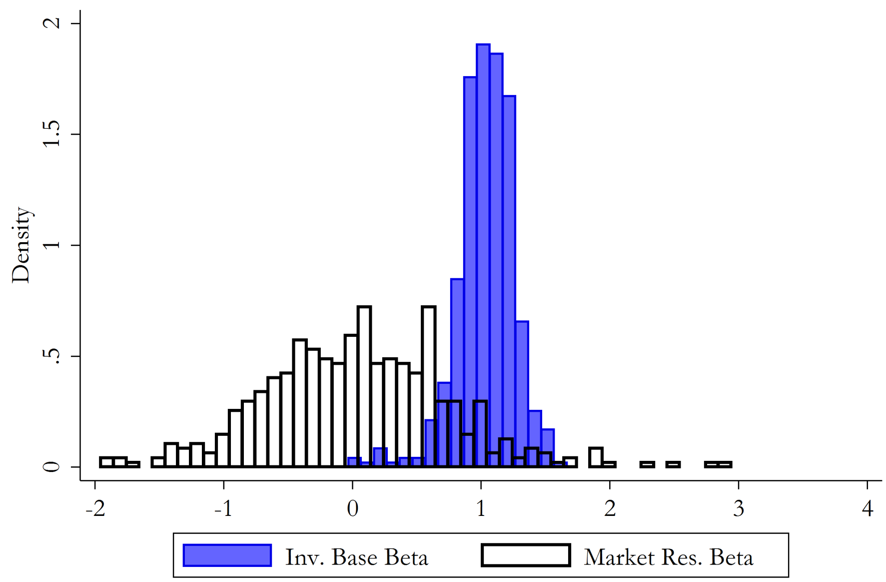


Figure IA2. Distribution of Currencies Market Beta and Connected Portfolio Beta. This figure shows the distribution of currency betas relative to market returns and connected portfolio returns. The market returns are the volume-weighted returns of all cryptocurrencies. For currency i, the connected portfolio returns are volume-weighted average returns of all the other currencies that are exposed to currency i, where a higher weight is assigned to currencies that have a higher trading volume and are more connected to i. To calculate the betas, first the market returns and connected portfolio returns are orthogonalized. Then 24-hour currency returns are regressed on the orthogonalized returns. Panel A shows the beta distributions when the market moves first and the connected portfolio returns are orthogonalized relative to the market, and Panel B shows when the connected portfolio returns move first and the market returns are orthogonalized.

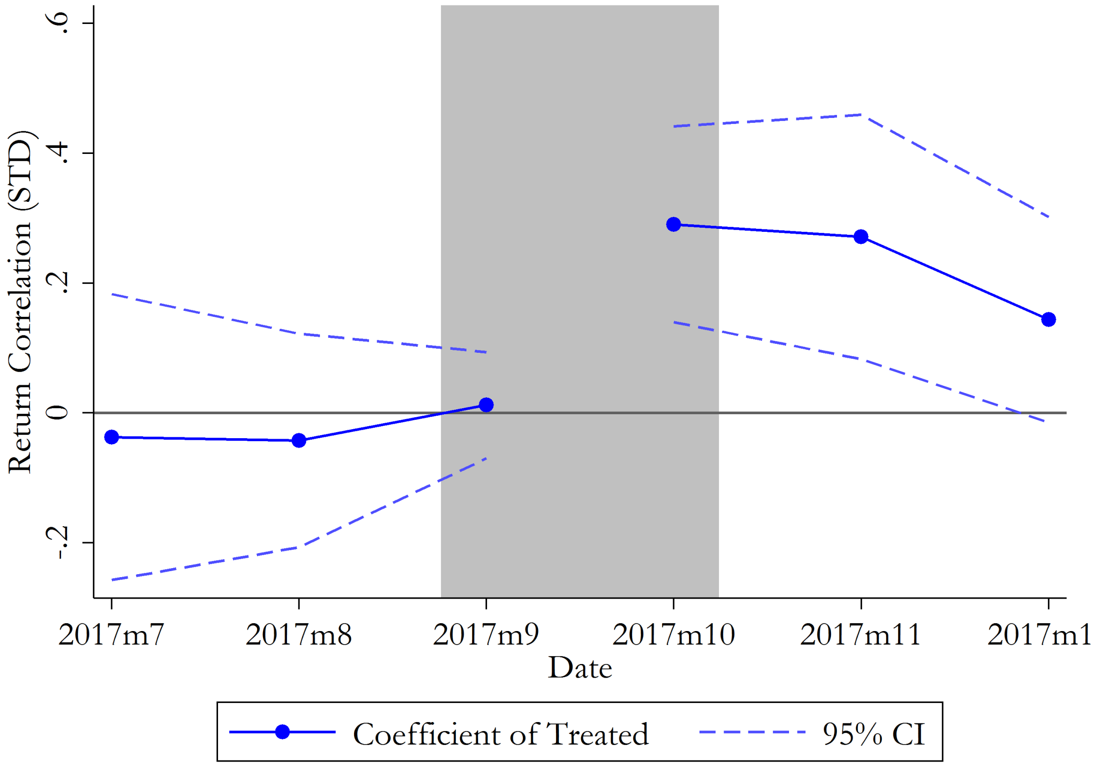


Figure IA3. Shutdown of Chinese Exchanges and Changes in Return Correlations. This figure shows relative changes in return correlations of cryptocurrency pairs whose connectivity was positively affected by the shutdown of Chinese exchanges. For each treated cryptocurrency pair, a control group is matched by finding the ten closest pairs based on the Mahalanobis distance in the pairwise connectivity and the monthly trading volume of both currencies the month before the shutdown. This figure shows the coefficient estimates of the standardized pairwise correlations on a treatment dummy controlling for similarity in volume, number of exchanges, and currency type. The blue dashed lines show the 95% confidence interval. The standard errors for estimation of the confidence intervals are clustered at the dyadic level.

Table IAII. Commonalities in Order Imbalance of Currencies on the Same Platform. This table reports the estimates for a panel regression similar to Table II for currencies that are listed on at least five exchanges and exchanges with at least five listed currencies. Panel A reports the results for the order imbalance based on the number of transactions ($OIBNUM$) and Panel B based on the dollar volume ($OIBVOL$). Standard errors are two-way clustered by currency and exchange. t-statistics are reported in parentheses. *p<.05, **p<.01, ***p<.001.

B. Order Imbalance Based on the Number of Transactions

	(1)	(2)	(3)	(4)	(5)
$\text{OIBNUM}^{\text{Cur}}$	0.171**			0.143*	0.166**
	(3.18)			(2.86)	(3.13)
$\text{OIBNUM}^{\text{Exch}}$		0.266***		0.254***	0.253***
		(6.08)		(5.09)	(5.02)
$\text{OIBNUM}^{\text{Mkt}}$			0.0764	-0.0246	-0.0295
			(1.76)	(-0.64)	(-0.77)
$\text{L.OIBNUM}^{\text{Cur}}$					-0.0249
					(-1.84)
$\text{F.OIBNUM}^{\text{Cur}}$					-0.0341***
					(-4.96)
Curr-Exch FE	Yes	Yes	Yes	Yes	Yes
Observations	37848	37848	37848	37848	37319
R^2	0.045	0.089	0.024	0.105	0.108
Curr & Exch Cluster	Yes	Yes	Yes	Yes	Yes

A. Order Imbalance Based on the Dollar Volume

	(1)	(2)	(3)	(4)	(5)
$\text{OIBVOL}^{\text{Cur}}$	0.238***			0.192**	0.226***
	(4.61)			(3.64)	(4.33)
$\text{OIBVOL}^{\text{Exch}}$		0.287***		0.258***	0.255***
		(6.24)		(4.73)	(4.74)
$\text{OIBVOL}^{\text{Mkt}}$			0.183***	-0.0152	-0.0193
			(5.88)	(-0.50)	(-0.65)
$\text{L.OIBVOL}^{\text{Cur}}$					-0.0381***
					(-4.12)
$\text{F.OIBVOL}^{\text{Cur}}$					-0.0450***
					(-4.73)
Curr-Exch FE	Yes	Yes	Yes	Yes	Yes
Observations	37843	37843	37843	37843	37313
R^2	0.059	0.088	0.038	0.118	0.123
Curr & Exch Cluster	Yes	Yes	Yes	Yes	Yes

Table IAIII. Exposure to the Same Investor Base and Comovement in Returns (Within Exchanges). This table reports estimates of a panel regression where the dependent variable is the within-month correlation of market model residuals between all cryptocurrency pairs that are listed on the same exchange. Exchange-level prices are used to calculate the return correlations. The market return is calculated as the volume-weighted returns of all cryptocurrencies, where each currency is weighted by its past 24-hour trade volume. Market residuals are estimated using a recursive time-series regression of individual currency returns on the market return. The $Connectivity$ measure is constructed as described in Section IV. $SameVol$ is the negative of the absolute difference in the percentile rank-transformed of monthly volume for each currency pair. $SameCoinToken$ is a dummy variable that takes the value of one if both currencies are coins or both tokens. Return correlations, the $Connectivity$ measure, and $SameVol$ are standardized by subtracting the mean and dividing by the standard deviation. t-statistics are reported in parentheses. *p<.05, **p<.01, ***p<.001.

	All Currencies			Large Currencies		
	(1)	(2)	(3)	(4)	(5)	(6)
Connectivity	0.165***	0.144***	0.142***	0.226***	0.192***	0.190***
	(9.19)	(8.77)	(8.67)	(11.57)	(9.07)	(9.05)
Similarity$^{\text{Volume}}$		0.040***	0.038***		0.149***	0.138***
		(3.94)	(3.74)		(3.63)	(3.34)
Similarity$^{\text{SameNExch}}$		0.101*	0.101*		0.111*	0.112*
		(2.58)	(2.58)		(2.24)	(2.26)
Similarity$^{\text{SameCoinToken}}$			0.045**			0.044*
			(3.29)			(2.24)
Exchange-Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	251522	251522	251522	81868	81868	81868
Adjusted R^2	0.218	0.229	0.229	0.192	0.211	0.211
Double Cluster	Yes	Yes	Yes	Yes	Yes	Yes