Stanford | Institute for Economic Policy Research (SIEPR)

Bitcoin Pricing, Adoption, and Usage: Theory and Evidence

Susan Athey

Ivo Parashkevov

Vishnu Sarukkai

Jing Xia

August, 2016

Working Paper No. 17-033

Bitcoin Pricing, Adoption, and Usage: Theory and Evidence *

Susan Athey[†] Ivo Parashkevov[‡] Vishnu Sarukkai[§] Jing Xia[¶]

First Draft: October 2013
This Draft: August 2016

Abstract

This paper develops a model of user adoption and use of virtual currency (such as Bitcoin), and focusing on the dynamics of adoption in the presence of frictions arising from exchange rate uncertainty. The theoretical model can be used to analyze how market fundamentals determine the exchange rate of fiat currency to Bitcoin. Empirical evidence from Bitcoin prices and utilization provides mixed evidence about the ability of the model to explain prices. Further analysis of the history of all individual transactions on Bitcoin's public ledger establishes patterns of adoption and utilization across user types, transaction type, and geography. We show that as of mid-2015, active usage was not growing quickly, and that investors and infrequent users held the majority of Bitcoins. We document the extent to which the attributes of the anonymous users of Bitcoin can be inferred through their behavior, and we find that users who engage in illegal activity are more likely to try to protect their financial privacy.

1 Introduction

The "digital currency" or "crypto-currency" Bitcoin [Nakamoto, 2008] has attracted great attention in the public press since early 2013, when its transaction volume and market capitalization became material. The goal of this paper is to explore Bitcoin theoretically and empirically from the perspective of the market for Bitcoin, its price determination, and its usage.

*We are grateful for comments provided at seminars at the ACM Electronic Commerce Conference (EC 2015), the Toulouse Network for Information Technology, Microsoft Research, Cornell University, University of Maryland Cybersecurity Workshop, and Stanford. Part of this research was conducted while Susan Athey was visiting Microsoft Research. Gleb Romanyuk provided subsantial and exceptional research assistance on the theoretical part of the model, while Christian Perez, Michael Piccirilli, and Phil Casey provided outstanding research assistance on the empirical work. We are grateful to Charles Songhurst for introducing the topic and Zachary Apter for early discussions about the exchange rate, as well as Dennis Carlton, Eric Budish, Bob Hall, Tom Sargent, and numerous participants in the Bitcoin industry. Meiklejohn et al. [2013] graciously shared some of their data with us.

[†]Graduate School of Business, Stanford University, and NBER. athey@stanford.edu
[‡]Charlie Finance. parashkevov@gmail.com
[§]Stanford University. sarukkai@stanford.edu
[¶]Cornerstone Research. Jing Xia <jing.xia16@gmail.com

Even defining Bitcoin is complex: it can be described as a protocol, a currency, a payment system, and a technology platform. At its core, Bitcoin is open source software, and services around Bitcoin are added on by a range independent companies and software developers. The software enables a public ledger of transactions, coupled with protocols and software that maintain security. An individual "owns" a Bitcoin if there is a ledger entry moving the Bitcoin to an address belonging to the individual; if the individual has the appropriate passcode, then the individual can in turn authorize a ledger entry assigning it to another individual's address. Unlike a bank balance that can be viewed or manipulated digitially, an individual's Bitcoin balance is not an "IOU" or a promise to provide funds on demand; the individual with the passcode associated with an address has full control over its disposition, and that Bitcoin balance is not linked to anything else. The full definition of Bitcoin ownership is the entry on the ledger assigning it to an address.

When an asset is purely digital, and when a public ledger can securely track ownership of assets, many new uses become possible. Recently, a variety of startups have used the Bitcoin ledger or Bitcoin-like technology as a platform to provide a variety of services, such as settlement of securities (NASDAQ is working on such an application [UKG, 2015]) or inter-bank settlement of fiat currency (Ripple reports that 10 of the top 50 banks have an active integration for payments and settlement [Elison, 2016]). In this paper we will not focus on these applications or the idea of Bitcoin as a technology platform, but instead consider direct consumer use of the technology, for example as a store of value or as a payment system.

For payments, Bitcoin has several distinguishing features. One individual can transfer Bitcoins to another anywhere in the world, without relying on counterparties or trust relationships beyond the trust in the software, and indeed without getting authorization to do so from any company or government. All that is necessary to transfer value electronically is the passcode to authorize a ledger entry to move the Bitcoins to another address. On the other hand, if the recipient wants to obtain value from the Bitcoins in the short term, she must either find a merchant who will accept them, find an individual buyer for the Bitcoins, or sell them on an exchange. This creates frictions and risk in using Bitcoin; these will be a focus in this paper.[1]

Given that in order to use Bitcoin as a payment system, users must be exposed to exchange rate fluctuations, the exchange rate of fiat currency to Bitcoin takes a central role in analyses of its utility. Indeed, fluctuating Bitcoin prices have attracted media attention and have been associated with billions of dollars of trading volume. There are self-service, internet-based currency exchanges that trade Bitcoin for fiat currency. Individuals may wire or otherwise transfer fiat currency to and from the exchanges, and send or receive Bitcoin electronically in exchange. The exchange rate for Bitcoin to fiat currencies floats and is determined by supply and demand; there is no governmental

[1]Some other unique features include the "push" (as opposed to "pull") nature of payments: the sender authorizes the movement, and the recipient does not require any information from the sender. Indeed, if the sender has the recipient's Bitcoin address, the recipient cannot block a payment even if she wants to. Consumer protection is also lacking: if a sender accidentally sends Bitcoins to the wrong address, they are gone, and nothing can be done. Similarly if a Bitcoin owner loses her address or passcode information. Some start-up firms have stepped in to provide additional layers of consumer protection, at the expense of requiring the consumer to trust the firms to stay in business, maintain security, etc.

authority or company making guarantees about its value. The supply of Bitcoins is essentially pre-determined, and they are created at a fixed rate over time up to a maximum. We leave a description of the creation of Bitcoins ("mining") and the maintenance of the secure ledger to others (see, e.g., Böhme et al. [2015] or Bonneau et al. [2015]), although there are fascinating economic issues involved in those aspects of Bitcoin as well.

The fact that Bitcoin is not backed by governments and also has no fundamental value has created confusion in the popular discussion about how exchange rates are determined. The first contribution of this paper is to show that Bitcoin exchange rates can (in a very stylized model) be fully determined by two market fundamentals: the steady state transaction volume of Bitcoin when used for payments, and the evolution of beliefs about the likelihood that the technology survives. In steady state, exchange rates are determined by the ratio of transaction volume to the supply of Bitcoins. In the absence of sufficient participation by investors (that is, when the demand for Bitcoin is primarily users of Bitcoin), the exchange rate also depends on the rate of adoption and the level of demand. In our model, exchange rates increase over time until they reach a steady state where all relevant users use the technology when they have a need that fits Bitcoin's use cases.

We then proceed to two distinct types of empirical work. The first examines the relationship between transaction volume and exchange rates over time, showing that exchange rates generally follow a pattern consistent with fundamentals outside a few (non-trivial) periods of price spikes. Thus, at least directionally, the forces of supply and demand appear to be operating.

Second, we look at individual behavior on Bitcoin to understand how Bitcoin is being used (in particular, whether for investment or for payments). This is possible because the ledger is public, although Bitcoin is "pseudonymous," meaning that identity information is not attached to addresses, addresses can be followed over time. However, since individuals typically have multiple addresses (due to some details of how addresses are managed, in "wallets"), it is not trivial to identify which collections of addresses are individual users or Bitcoin businesses. Building on Ron and Shamir [2013] and Meiklejohn et al. [2013], we implement a variety of heuristics, in combination with publicly available data sources, to group addresses into entities and identify businesses. In all, we use a list of approximately 220,000 entities whose addresses are known to belong to businesses with high probability, and we take about 78 million Bitcoin addresses and group them into 27 million distinct "entities."

Using this data, we then analyze user adoption and behavior. We show that Bitcoin ownership is highly concentrated, and that only a small fraction of users fall into the category of long-term, frequent users (where frequent users transact more than 10% of days in their time as active users). Most users only do a small number of transactions. Frequent use is not growing as a share of the network. A substantial share of transaction volume takes place with exchanges. Thus, investors likely play a major role in transaction volume at this time. Viewed through the lens of our theoretical model, this implies that exchange rates are likely to be more sensitive to beliefs about the future of Bitcoin and less sensitive to current transaction volume for its own sake (though current volume is likely a signal of the future).

We then examine several use cases of Bitcoin in more detail. One use of Bitcoin that has gained public attention is the use of Bitcoin to buy illegal substances such as drugs and firearms. The website Silk Road famously operated as a reputation-based peer-to-peer marketplace in drugs that used Bitcoin for payments, until it was closed. We can only identify about 1% of the dollar value of transactions as relating to contraband, and a similarly small share for gambling, despite having addresses in our data for the largest known entities in these categories (for example, we identify contraband sites Silk Road, Silk Road 2, AgoraMarket, and Evolution Market, together accounting for $9 Billion in transactions, as well as gambling site Satoshi Dice, which accounts for $2 billion). We also look at the attempts of users to protect their privacy. There is a system called CoinJoin that mixes balances from different addresses and sends them to new addresses, in an attempt to disguise the origin of the funds. The use of CoinJoin of course complicates our attempts to analyze user behavior, and our heuristics for grouping addresses into entities adjust for the use of CoinJoin, while adding noise (making it hard to track CoinJoin users over time). We find that contraband buyers are more likely (19% of all buyers) than the general population (12%) to use CoinJoin to protect their privacy; gamblers, by contrast, are very similar to the general population, also having approximately 12% of all gamblers in Bitcoin history using Coinjoin. It still may be surprising that so many contraband buyers do not use this service.

Another key use case of Bitcoin is international payments. Given that we do not have identity information about the users, it is difficult to directly analyze international transfers. We attempt to classify users into four geographical regions using their behavior, based on a training dataset of 2858 users whose country of origin is known. We obtain a moderate classification rate (roughly 60%) using features such as the times they transact and the exchanges they use. This exercise also serves to illustrates that much can be learned about users even without directly having identifying information. Using the classifications, we analyze adoption curves and patterns of behavior by country. We see that transaction volume is more concentrated within region, but there is a fair amount of cross-region trade as well.

Overall, the micro-data support the conclusion that the most important use case for Bitcoin today is investing (store of value); this makes it harder to link exchange rates to current fundamentals, but rather puts more weight on beliefs about the future. Despite the pseudonymous nature of the system, we can provide rich desciptive statistics about the use of Bitcoin, and we can classify users moderately well based on their behavior. All conclusions from micro-data are subject to the caveat that numerous heuristics were used in transforming the raw data into the entities.

2 Related Work: Bitcoin and other Digital Currencies

Although Bitcoin is only a few years old, it has already attracted attention in a variety of different disciplines. Regulatory issues have been analyzed by Brito and Castillo [2013] and Trautman [2014]. Several economists have described the institutions and policy issues surrounding Bitcoin, including Böhme et al. [2015], Evans [2014], and Van Alstyne [2014]. Gans and Halaburda [2013] examine

virtual currencies within a platform (e.g. in-game currency), while Gandal and Halaburda [2014] and Halaburda and Sarvary [2016] look at competition between digital currencies.

The volatility and predictability of prices have been considered by Dwyer [2015], Surda [2012], Kristoufek [2015], Gourieroux and Hencic [2014], Donier and Bouchaud [2015], Kaminski and Gloor [2014], and Ciaian et al. [2015], while Moore and Christin [2013] shows that the failure of Bitcoin exchanges can be predicted with transaction volume. We defer a discussion of the relevant macroeconmic literature to the next section.

The characteristics of Bitcoin users have been studied through surveys [Bohr and Bashir, 2014] and Google trends data [Yelowitz and Wilson, 2015]. One of the largest single use cases, purchasing drugs and other contraband, received an in-depth analysis by Christin [2013], who scraped thousands of transactions from the website.

Another area that has attracted attention in the computer science community is the extent to which user transactions are private or anonymous, can be identified using approaches such as associating internet protocol addresses with activity (see Androulaki et al. [2013], Koshy et al. [2014], Biryukov et al. [2014]), or can be hidden using alternative protocols or "mixing" (e.g., Androulaki and Karame [2014], Meiklejohn and Orlandi [2015]). Reid and Harrigan [2013] and Ron and Shamir [2014] show that thefts or other large-scale illegal uses of Bitcoins can be analyzed and understood from the public ledger.

Closer to the spirit of the current paper, Ron and Shamir [2013] introduced a variety of heuristics that could be used to combine Bitcoin addresses into "wallets", and presented summary statistics about the distribution of Bitcoin holdings and transactions. Meiklejohn et al. [2013] built on the entity creation approach, and further collected a large number of Bitcoin addresses for a variety of entities including exchanges, contraband, gambling, and others. We make use of this list of known addresses in our analylsis. Ober et al. [2013] and Kondor et al. [2014] use a similar approach to document properties of the network, including summary statistics of network constructs such as the degree distribution.

Other issues that have received attention include the game theoretic and incentive issues around mining and security (e.g. Karame et al. [2012], Babaioff et al. [2012], Kroll et al. [2013], and Donet et al. [2014]).

3 A Theoretical Model of Bitcoin Exchange Rates

This section builds a theoretical model of Bitcoin adoption and the determination of Bitcoin-to-dollar exchange rates, with the goal of showing that there exists a coherent set of assumptions under which Bitcoin exchange rates are fully determined by economic primitives. We begin by summarizing some of the economic forces at work in the real-world Bitcoin market, and then build a model that captures a few key forces (abstracting away from many others).

Bitcoin shares some features with "private money"–money that is issued by a private party rather than a central government. Some examples of private money include local currency that is usable

within a network of merchants (e.g. "Ithaca hours" or BerkShares), notes issued by private banks whose value is tied to a government-backed currency, and notes issued by banks whose value floats. In many cases there is an element of uncertainty in the future value of private currencies (in the U.S. this uncertainty was addressed through regulation requiring private banks to hold U.S. government bonds to back the notes). Private money has been criticized along a variety of fronts, including the exchange rate fluctuations and associated inefficiencies it brings, the lack of control over the money supply that it engenders, and the uncertainty due to the possibility of bank runs ([Fischer, 1995]). Bitcoin does not have the same issues that arise in fractional reserve banking, and so it can be thought of as private money with some unique characteristics, such as immunity from bank runs. It may bring other risks, such as technological risk.

One strand of literature that is more directly relevant for Bitcoin looks at the ability of private currency to exist in an environment where currency issuers can commit to future issuance of currency; in a world with positive real interest rates, in order for agents to hold the currency, it needs to be deflationary in nature, requiring the issuer to commit to buy back currency over time ([Marimon et al., 2012]). Bitcoin has precommitted to a total number of Bitcoins ever to be issued; in the meantime, Bitcoins continue to be issued at a rate that is essentially prespecified, while seignorage is used to compensate individuals who participate in maintaining the transaction ledger. Thus, the Bitcoin protocol addresses issues of commitment.

However, since Bitcoin is not backed by an underlying asset and instead has a fully fluctating exchange rate, there is substantial risk about its future value. If Bitcoin has a functional use as a currency or a medium of transferring money, then its future value may be tied to the future magnitude of that use. In these regards, it also shares some features with risky assets where beliefs about the future underlying value of the asset evolve over time as information is revealed. In such environments, speculative "bubbles" can form, and there can be room for conflicting beliefs by investors. Given many of the fluctuations that have occured with Bitcoin exchange rates, the idea of bubbles seems salient for Bitcoin.

More broadly, another issue that has received attention in the macroeconomics literature is the idea that "money is memory" ([Kocherlakota, 1998]). Indeed, this economics literature received attention from early Bitcoin industry participants, and scholars also observed that Bitcoin (as a purely digital recordkeeping device) fits neatly into this theoretical framework ([Luther and Olson, 2013]). As technology lowers the costs of public recordkeeping devices, it seems natural that a digital ledger might emerge as an alternative to physical coins.

In this paper we build a model that incorporates some but not all of these issues. In our model, the exchange rate for Bitcoin is determined through supply and demand by users, and in an extension, investors. User demand is based on the ability to make international payments, where alternative methods have high fees. Bitcoin entails a cost of effort as well as exchange rate risk. The exchange rate risk is endogenously determined and changes over time. There is uncertainty about the underlying value of the technology; we incorporate this as an unobserved state of the of technology, where if there are defects in the technology, Bitcoin may "break down" leading to a loss

of the coins. With heterogeneous effort costs and continually improving beliefs about the quality of the underlying technology, in our model adoption grows over time, with the steady state determined by the usage of Bitcoin. Investors accelerate the evolution of the exchange rate path relative to users, whose risk aversion and effort costs can lead to opportunties for investor purchases of Bitcoin.

3.1 Model Setup

This section develops the theoretical model that we will use to understand the value of Bitcoin and the gradual adoption of Bitcoin in the population. The model illustrates that the exchange rate for Bitcoin can be grounded in economic fundamentals, and that the exchange rate rises with usage.

We consider a model where the Bitcoin's primary use is to transfer money across borders–a remittance model. Similar forces arise in a model where Bitcoin is used for payments, with a few additional complications (such as the choice of how the gains from Bitcoin usage are shared between buyer and seller).

Our model leaves out a number of important factors. We do not incorporate heterogeneous beliefs or speculative bubbles. We also do not incorporate any state variables on the user side other than beliefs; in particular, we do not account for network effects in terms of awareness or utilization. We have a fixed number of agents who can in principle use Bitcoin in each period, and once all agents have adopted, the exchange rate fluctuates around a steady state in response to demand shocks. Our model thus leaves out the possibility that there are always agents at the margin for adoption, and it does not fully incorporate an important real-world effect, that volatility of Bitcoin can decline substantially as the market thickens. Another related factor that is left out is the competitive response of other firms: other virtual currencies might enter the market, or banks may cut fees (or adopt technology that reduces their costs and delays for payments).

3.1.1 Technology and Beliefs

Bitcoin is a new technology which may or may not have defects. The unknown quality of Bitcoin is $S \in \{0, 1\}$. The state is $S = 1$ ("good") if Bitcoin is a functional technology. The state is $S = 0$ ("bad") if Bitcoin has a defect. Let q_t be the public belief at time t that BC is good:

$$q_t := Pr_t(S = 1).$$

At the end of each period, Bitcoin may "break" so that all ongoing transactions are lost. If the state is good, Bitcoin survives forever; if the state is bad, Bitcoin falls apart with probability λ:

$$\begin{aligned} \Pr(\text{break}|S = 0) &= \lambda \in (0, 1) \\ \Pr(\text{break}|S = 1) &= 0 \end{aligned}$$

The probability that BC survives until time $t+1$ given that it survived until time t is then equal to $\kappa_t \equiv q_t \cdot 1 + (1 - q_t)(1 - \lambda)$. The initial prior about the state is commonly known and is denoted q_0.

The evolution of beliefs about S is straightforward. If Bitcoin falls apart before time t, $q_t = 0$. If BC survives until time t, q_t is updated upwards by Bayes' rule:

$$\begin{aligned} q_t &= \frac{q_{t-1}}{q_{t-1} + (1-\lambda)(1-q_{t-1})} \\ &= \frac{q_0}{q_0 + (1-\lambda)^t(1-q_0)}. \end{aligned}$$

Conditional on $S = 1$, q_t increases monotonically with t, approaching 1 as t grows large.

3.1.2 Users

There are $i = 1, .., N$ potential users of Bitcoin. Their use case is remittances. There is a time cost of using Bitcoin; each individual's cost is given by $c_i \in [\underline{c}, \bar{c}]$, with $\underline{c} \geq 0$, which does not vary over time. For simplicity we assume that $c_i \neq c_i'$ for all $i, i' \in \{1, .., N\}^2$.

In each period t, each agent i needs to send a random amount x_{it} to his family. We assume that x_{it} is distributed Bernoulli, taking the value 1 with probability p. We use $\mathbb{E}_{i,t}$ to denote the expectation over random variables using player i's beliefs given information available at the beginning of period t. We let $\tilde{e}_{t+1}$ be the random variable whose realization is the exchange rate in time t conditional on Bitcoin surviving. Denote the distribution of $\tilde{e}_{t+1}$ from period t's perspective by G_{t+1}.

In the beginning of period t, every agent observes his need x_{it}. Then exchange rate e_t realizes and each agent chooses action $a_{it} \in \{B, D\}$: send x_{it} via Bitcoin or or send x_{it} via bank transfer. Lastly, Bitcoin's survival is observed. If Bitcoin survives, the relatives receive the Bitcoins and exchange them for dollars at rate e_{t+1}; if Bitcoin breaks, the relatives receive zero. The agent's preferences exhibit risk aversion and are represented by an increasing and weakly concave function v.

Player i's payoff from using Bitcoin at time t is then written:

$$u(B; x_{it}, e, \kappa_t, c_i) = \kappa_t \cdot \mathbb{E}_{it} v\left(\frac{\tilde{e}_{t+1}}{e} x_{it}\right) - c_i.$$

The banking system charges an upfront, fixed transfer fee f, and there is no risk of using the banking system. Thus the utility of using the banking system is written:

$$u(D; x_{it}, e, \kappa_t, c_i) = v(x_{it} - f). \tag{1}$$

We make a simplifying assumption on parameters that guarantees that full participation is possible when t is large. Let $\tilde{Z}$ be a random variable distributed $Bi(N, p)$. The assumption is then:

$$\mathbb{E} v\left(\frac{\tilde{Z}}{N}\right) - \bar{c} > v(1-f) > 0. \tag{2}$$

3.2 Equilibrium

For each realization of transfer needs $x_t = (x_{1t}, ..., x_{Nt})$, the *aggregate dollar demand curve for Bitcoins* in period t is

$$Q_t(e; x_t) = \sum_i x_{it} 1\{a_{it} = B\} = \sum_i x_{it} 1\{u(B; x_{it}, e, \kappa_t, c_i) \geq u(D; x_{it}, e, \kappa_t, c_i)\}. \tag{3}$$

The supply is fixed at $\bar{B}$.

Due to the finite number of agents in our model and our simplifying assumption of unit demand for each, aggregate demand for Bitcoins in each period is a step function. We need to determine how to clear the market in case it does not clear exactly. To keep things as simple as possible, we assume the existence of a market maker with unlimited funds who will buy between 0 and 1 Bitcoins in each period at the highest exchange rate where supply is greater than demand; the market maker will then sell them at the market price in the next period (which will be the highest exchange rate where supply is greater than demand in that period). Also to keep things simple, we assume in the event that there is no demand for Bitcoins, this market maker will buy all the Bitcoins at a very small fixed price. We do not incorporate the formalism for this in the model as it occurs with very small probability once N is sizeable.

The dynamic price equilibrium is given by specifying in each period t: (i) the exchange rate e_t, (ii) strategy of each agent $\sigma_{it}\colon \{0,1\} \times \mathbb{R} \mapsto \{0, B, D\}$, and (iii) each agent's belief $\mu_{it} \in \Delta(\mathbb{R})$ about tomorrow's e_{t+1}.

Profile $\{e_t, \{\sigma_{it}, \mu_{it}\}_i\}_t$ constitutes a dynamic price equilibrium if:

1. (Agent optimization) for every agent i, σ_{it} is optimal given e_t and x_{it}:

   $$\sigma_{it}(x_{it}, e_t) \in \arg\max_{a \in \{B,D\}} u(a; x_{i,t}, e_t, \kappa_t, c_i) \quad \forall x_{it}, e_t,$$

   and further we require that if the agent is indifferent between several choices and B is optimal, B is selected.

2. (Market clearing) for any realization of $(x_{1t}, ..., x_{Nt})$, e_t is the minimal exchange rate such that demand is less than supply:

   $$e_t = \min\{e : Q_t(e; x_t) \leq \bar{B}e\}. \tag{4}$$

3. (Rational expectations) Agent's expectation of next period's exchange rate conditional on Bitcoin surviving to the next period is correct:

   $$\mu_{it} = G_{t+1}.$$

Theorem 1. *(i) A dynamic price equilibrium exists and is unique. Denote the equilibrium exchange rate function as $e^*(x_t)$, and let the equilibrium agent strategies be σ^*_{it}.*

(ii) *Let $C_t^*(x_t) = max_i\{c_i : \sigma^*(1, e_t^*(x_t)) = B\}$ denote the highest fixed cost that adopts Bitcoin when they have a need in period t, given x_t. $C_t^*(x_t)$ is increasing in t, and thus G_{t+1} FOSD G_t (weakly).*

(iii) *There exists a t^c such that for all $t' \geq t^c$, all agents use Bitcoin for remittances ($\sigma_{it'}^*(1, e_t^*(x_t)) = B$ for all x_t), while for each $t'' < t^c$, there exists a realization of remittance needs $x_{t''} \in \{0,1\}^N$ such that for some i, ($\sigma_{it''}^*(1, e_t^*(x_t'')) = 0$).*

Proof. First, we show that (2) implies that such a $\hat{t}$ exists whereby all agents participate in Bitcoin for $t > \hat{t}$. Since $q_t \to 1$, assumption (2) implies that there is $\hat{t}$ such that

$$(q_{\hat{t}} + (1 - q_{\hat{t}})(1 - \lambda))\mathbb{E}v\left(\frac{\tilde{Z}}{N}\right) - \bar{c} > v(1 - f).$$

We wish to show that for $t > \hat{t}$, it is a best response for all players to choose Bitcoin, given that they believe all agents will do so in the future. Note that if all players use Bitcoin in a given period t, the exchange rate is approximately (due to integer constraints) equal to $e_t = \sum_i x_{i,t}/\bar{B}$. (Note that the approximation is due to integer constraints in market clearing; we assume that N is large enough that the approximation error is small enough so that the inequality still holds; alternatively we could strengthen the assumption on primitives (2). We omit these details for expositional simplicity.) The payoff for agent i from using Bitcoin when $x_{i,t} = 1$ and he believes players in $t + 1$ use Bitcoin is

$$\begin{aligned} u(B; 1, e_t, \kappa_t, c_i) &= \kappa_t \mathbb{E}_{it} v\left(\frac{\tilde{e}_{t+1}}{e_t}\right) - c_i \geq \kappa_t \mathbb{E}_{it} v\left(\frac{\tilde{e}_{t+1}}{N/\bar{B}}\right) - \bar{c} \\ &\approx \kappa_t \mathbb{E} v\left(\frac{\tilde{Z}}{N}\right) - \bar{c} > v(1 - f) = u(D; 1, e_t, \kappa_t, c_i). \end{aligned}$$

Second, we use this to show that we can solve for a unique G_t in each t. Recall that G_t is the distribution of $\tilde{e}_{t+1}$. Now, observe that at time $\hat{t}$, there is full participation, and so $\tilde{e}_{\hat{t}+1}$ is known (and approximately equal to the distribution of $\tilde{Z}/\bar{B}$). The inductive step: if we assume that G_{t+1} is known, we find G_t as follows. In period t, rational expectations imply that $\mu_{it} = G_{t+1}$ for all i. Therefore, for every x_t the aggregate demand $Q_t(e; x_t)$ in (3) is well-defined, continuous from the left and decreases from $\sum_i x_{it}$ to 0 as e increases from 0. Therefore, there is a unique market-clearing exchange rate e_t that solves (4). The distribution of x_t then induces a distribution over e_t. By induction, there is a unique G_t for all t.

Third, we show that $C_t^*(x_t)$ is (weakly) increasing in t. In period $\hat{t}$, all users will use Bitcoin if they need a transfer. In period $\hat{t}$, due to the probability that Bitcoin breaks down, for all e, the expected payoff to each player i from Bitcoin is uniformly lower than if it were period $\hat{t} + 1$ (noting that beliefs are stationary after period $\hat{t}$):

$$u(B; 1, e, \kappa_t, c_i) = \kappa_{\hat{t}} \cdot \mathbb{E}_{i,\hat{t}} v\left(\frac{\tilde{e}_{\hat{t}+1}}{e}\right) - c_i \tag{5}$$

$$< \kappa_{\hat{t}+1} \cdot \mathbb{E}_{i,\hat{t}+2}\left[v\left(\frac{\tilde{e}_{\hat{t}+2}}{e}\right)\right] - c_i = \kappa_{\hat{t}+1} \cdot \mathbb{E}_{i,\hat{t}+1}\left[v\left(\frac{\tilde{e}_{\hat{t}+1}}{e}\right)\right] - c_i.$$

By symmetry of the agents, this in turn implies that the set of cost realizations that will select Bitcoin is higher in period $\hat{t}+1$ than $\hat{t}$, and further, for any $x \in \{0,1\}^N$ and any e, $Q_{\hat{t}}(e;x) \leq Q_{\hat{t}+1}(e;x)$, that is, the demand curve is lower at $\hat{t}$ than at $\hat{t}+1$ for any realization of demand shocks. This in turn implies that $G_{\hat{t}+1}$ FOSD $G_{\hat{t}}$.

We then complete the proof of part (ii) using induction. Suppose that G_{t+1} FOSD G_t for some t. Then we have (using $\mathbb{E}_G$ to denote the expectation over an exchange rate with distribution G) that

$$u(B;1,e,\kappa_{t-1},c_i) = \kappa_{t-1} \cdot \mathbb{E}_{G_t} v\left(\frac{\tilde{e}_t}{e}\right) - c_i \leq \ \kappa_t \cdot \mathbb{E}_{G_{t+1}} v\left(\frac{\tilde{e}_{t+1}}{e}\right) - c_i = u(B;1,e,\kappa_t,c_i). \tag{6}$$

Therefore, for any realization of $x \in \{0,1\}^N$, the demand curve in period $t-1$ is lower than the demand curve in period t: $Q_{t-1}(e;x) \leq Q_t(e;x)$. This implies that $C_t^*(x) \leq C_{t+1}^*$, and further G_t FOSD G_{t-1}. The result then follows by induction.

Finally we complete the proof of part (iii). Given that $C_t^*(x)$ is increasing in t, together with the fact that there exists a $\hat{t}$ where all agents use Bitcoin for all x, it follows that there is a unique lowest t where all agents use Bitcoin for all x. This is denoted t^c as in part (iii).

□

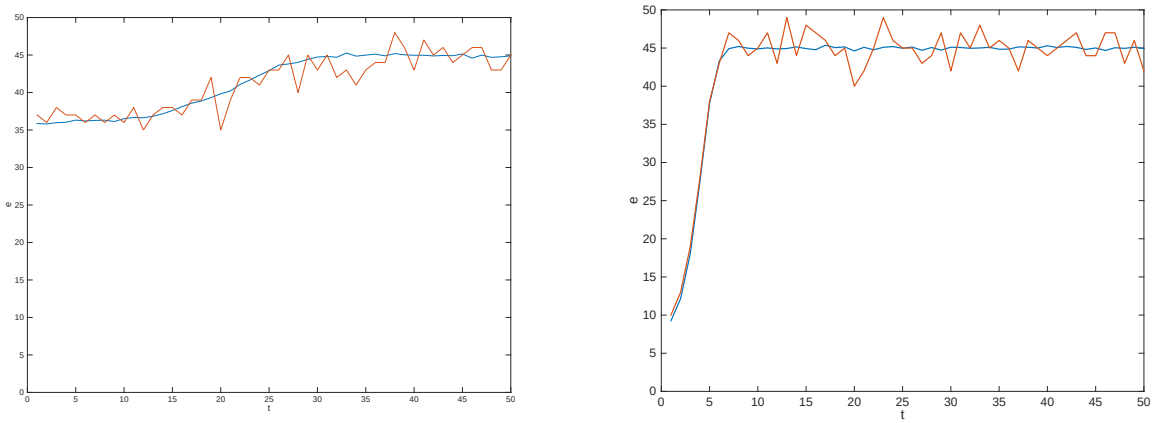
Our uniqueness result depends on a number of assumptions. The ultimate complete adoption assumption is important. Since the units of Bitcoin are arbitrary, it would always be possible for an investor to exogenously buy Bitcoins and take them out of the market, in which case the value of remaining Bitcoins would increase proportionally. We also impose the simplifying assumption that investors stop investing once a steady state is reached.

3.3 Dynamics of Adoption Rate

Theorem 1 establishes that adoption increases over time, eventually reaching full adoption. The adoption path depends on a number of factors, including the prior probability of the bad state, the probability of failure in the bad state, the risk aversion of the agents, and the number of agents.

Figure 1 illustrates two paths for the exchange rate (which is proportional to adoption) for different values of λ, the probability of failure in the bad state. The higher value of λ leads to lower initial adoption rate but faster learning.

We now consider the role of risk in the dynamics. First consider the dynamics of risk itself. Early on, we expect that a relatively small number of agents will use Bitcoin, and so the realized demand for Bitcoin should fluctuate around its expectation. A counterveiling force in this model is that if some agents with low costs of using Bitcoin realize $x_{it} = 0$, other agents with higher costs may be attracted by the lower exchange rate. Once the critical time t^c has passed, all agents participate, and so if an unexpectedly low number have remittance needs, there are no additional higher-cost agents to fill the gap. Similarly, the probability of failure is low enough that even if all



(a) $\lambda = .2$ (b) $\lambda = .7$

Figure 1: Equilibrium distributions of exchange rate $\{G_t\}$. Blue line is $\{\mathbb{E}_t e_t\}$, red line is one simulated draw $\{e_t\}$. Parameter values: $N = 50$, $\bar{B} = 1$, $p = .9$, $v(x) = 1 - e^{-x}$, $q_0 = .005$, $f = .5$, $[\underline{c}, \bar{c}] = [0, .46]$.

agents have remittance needs, no agents are deterred from using Bitcoin. In reality, we expect that as a market grows "thicker" that exchange rate volatility should decrease, but this model does not incorporate all of the forces that would lead to this outcome.

We can highlight the fact that risk does play a deterrant role for adoption by comparing adoption paths when agents have higher levels of risk aversion. Figure 2 illustrates the impact of greater risk aversion: it slows the adoption path.

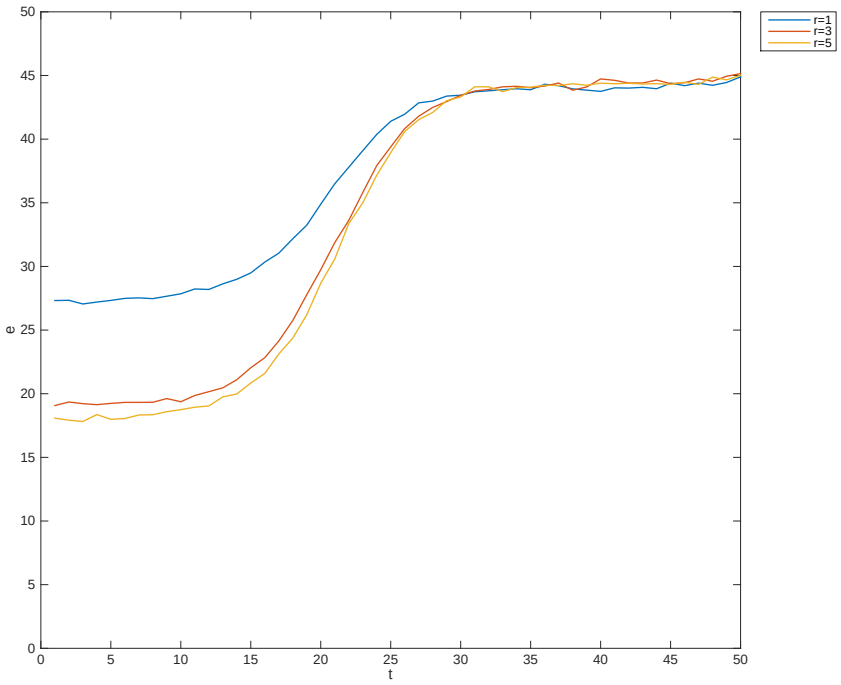


Figure 2: Equilibrium average exchange rates $\{\mathbb{E}_t e_t\}$ for different coefficients of risk aversion. Parameter values: $N = 50$, $\lambda = .3$, $\bar{B} = 1$, $p = .9$, $v(x) = 1 - e^{-rx}$, $q_0 = .005$, $f = .5$, $[\underline{c}, \bar{c}] = [0, .5]$, $r \in 1, 3, 5$.

3.4 Extension: Investors

A perhaps-unrealistic feature of our model is that due to the risk aversion and time costs of the remitting agents, the exchange rate might grow slowly relative to the evolution of beliefs. Indeed, Figure 2 shows that for the same q_t process, the exchange rate can start lower but then grow more quickly for a period of time with more risk averse users. In such a setting one might expect investors to enter in order to take advantage of the fast-growing exchange rates, buying Bitcoins simply because they are expected to gain sufficient value in the next period to offset the risk of Bitcoin breaking down.

An obvious but still important initial observation is that if the supply $\hat{B}$ of Bitcoin changes exogenously and permanently, the equilibrium exchange rate path adjusts so that supply equals demand with the new, limited supply; Bitcoin values are simply rescaled. This observation has important implications for multiplicity of equilibria: if an investor, for exogenous reasons, committed to buy a fixed amount of Bitcoin in each period (forever), the rest of the market would adjust as if those Bitcoins had never existed. In this section, we will focus on scenarios where investors eventually stop investing when the market reaches steady state.

We thus introduce a speculative investor with the following characteristics:

- risk-neutral;
- unconstrained liquidity;
- rational expectations;
- profit-maximizing;
- no discounting (patient).

A single investor may place a "market order" for $y_t < B$ Bitcoins prior to the realization of individual demands, and may also sell short y_t Bitcoins. (The single investor is intended to be a simple way of introducing a large number of speculators whose trades reach equilibrium, without providing all of the micro-foundations.) Short selling requires a more complex institutional environment, since Bitcoin is being used for a purpose here (transfering money digitally) and is not simply an asset; here for simplicity, we rule out short selling.

Note that the investor does not affect the supply of Bitcoin. In period t, investors and remitters buy the entire supply of Bitcoins $\bar{B}$. In the next period $t+1$, the receivers and investors sell their Bitcoins, therefore the supply is again $\bar{B}$.

In the presence of the investor, we replace (4) in the definition of dynamic price equilibrium with

$$Q_t(e_t; x_t) = (\bar{B} - y_t) \cdot e_t. \tag{7}$$

Define $\hat{e}(x_t, y_t)$ to be the level of the exchange rate that satisfies (7); this exists by the arguments of Theorem 1.

The investor will select $y_t = 0$ if

$$\mathbb{E}_{x_t}\big[\mathbb{E}_{t+1}\big[\frac{\kappa_t \cdot \tilde{e}_{t+1}}{\hat{e}(\tilde{x}_t, 0)}\big]\big] \leq 1, \tag{8}$$

and otherwise, y_t satisfies

$$\mathbb{E}_{x_t}\big[\mathbb{E}_{t+1}\big[\frac{\kappa_t \cdot \tilde{e}_{t+1}}{\hat{e}(\tilde{x}_t, y_t)}\big]\big] = 1. \tag{9}$$

We refer to a dynamic price equilibrium and a vector $\{y_t\}$ such that for each t, (7) and either (8) or (9) hold, as a dynamic price equilibrium with investment.

Beliefs will be modified by the presence of investors, as well. As discussed above, there are potential indeterminancies once we introduce investors, due to the fact that the market revalues Bitcoins if it expects them to be purchased by investors. In addition, after we have full adoption of Bitcoin, then to have positive investment in one period, the investor must expect higher exchange rates in the next period in the absence of the investment; but since usage is constant (and there is discounting due to κ), this is only possible if there is an increasing level of investment in each period (creating artificial scarcity that increases over time).

For simplicity we instead focus on the case where in periods t^c and greater, the investor stops investing. For $t < t^c$, the investor will find it optimal to invest or sell short if the adoption rate of Bitcoin is too slow relative to q_t and expected future adoption; in such cases, the investor will buy Bitcoins, increasing the period t exchange rate. This puts a lower bound on the expected growth rate of exchange rates in any period. Since the final expected exchange rate is fixed, we will see that the exchange rate path will go up relative to a model without investors; it will reflect the expectation of the final exchange rate (based on the transaction volume) discounted by the probability that Bitcoin falls apart. Formally:

Theorem 2. *In the presence of investors who can buy and sell short Bitcoin, there is a dynamic price equilibrium with investment where for $t \geq t^c$, $y_t = 0$ and the distribution (given period t information) of $\tilde{e}_{t+1}$ is equal to that of $\tilde{Z}/\bar{B}$. In each period $t < t^c$,*

$$\mathbb{E}_{\tilde{e}_t, \tilde{e}_{t+1}}\left[\kappa_t \cdot \frac{\tilde{e}_{t+1}}{\tilde{e}_t}\right] \leq 1. \tag{10}$$

Proof. Note that the sequence $\{y_t\}$ is predetermined and the investment level does not depend on any state variables other than which period it is, since there is no information available to the investor at time t that is relevant to predicting future exchange rates (other than the survival of Bitcoin, where the evolution of beliefs conditional on survival is also predetermined). If for $t \geq t^c$, the investor and agents expect that $y_{t'} = 0$ for all $t' > t$, then in period t, (8) holds, since when $y_t = 0$, all agents adopt Bitcoin in both t and $t+1$, and the distribution of exchange rates in period t is equal to that in period $t+1$ (conditional on survival). In period $t = t^c - 1$, given that adoption is less than complete for some realizations of x_t by Theorem 1, it is possible that (8) does not hold. If it does, then beliefs in period t are the same as in the equilibrium of Theorem 1. We continue to check (8) in successively lower values of t, where beliefs about $\tilde{e}_{t+1}$ are worsening in the FOSD order in each period by the arguments of Theorem 1. If a period is reached where (8) fails, then a positive value of y_t is selected in that period. We can solve backwards for values of y'_t for $t' < t$ using (8) and (9), with beliefs in period t' induced by the y''_t for $t'' > t'$ and the uncertainty over $x_{t'+1}$. □

Thus, the presence of investors substantially narrows down the exchange rate path, decoupling it in expectation from the primitives of the consumer adoption process, and instead linking it directly to two key primitives: Bitcoin utilization in steady state, and beliefs about the probability that Bitcoin survives period to period.

3.5 Summary of Theoretical Results

Our model thus delivers the following:

1. In the absence of investors, if all agents eventually adopt Bitcoin, then there is a unique equilibrium exchange rate in each period, which is determined by supply and demand (economic fundamentals).

2. The steady state expected exchange rate for Bitcoin is equal to the ratio of the expected transaction volume and the supply of Bitcoins.

3. Transaction volume in turn depends the advantages of Bitcoin relative to other payment options. The initial level of supply of Bitcoins is irrelevant since Bitcoins are infinitely divisible, though it is important that the total supply is exogenous.

4. Beliefs about the quality of Bitcoin evolve over time and lead to increasing adoption over time, conditional on survival.

5. Bitcoin utilization starts out lower when agents are more risk averse.

6. The nature of learning and beliefs also influence the evolution of adoption and exchange rates.

7. Investors may buy Bitcoin, which decreases the effective supply for users and increases the market equilibrium price.

8. The exchange rate increases proportionally to adoption in the absence of investors, while it starts at a higher level and then increases more gradually to the steady state in the presence of investors, while investors stop investing once steady state is reached.

9. If an investor exogenously bought a certain quantity of Bitcoins and held them (or alternatively, if some were lost), the exchange rate would adjust in proportion to the number taken out of the market.

3.6 From Theory to Empirics

The theoretical model suggests a number of empirical hypotheses to test and questions to answer:

1. How tightly correlated are prices and transaction volume for Bitcoin? Is price directly proportional to transaction volume divided by the number of Bitcoins?

2. What share of Bitcoin activity appears to be about using Bitcoin as a store of value, versus international payments or e-commerce?

3. What use cases appear to be prevalent among commercial uses? Is illicit activity important, and do users appear to protect their financial privacy?

For the first question, if Bitcoin prices are determined primarily by current demand for transactions and velocity is relatively constant over time, the theory suggests that prices should move in proportion to transaction volume. The answer to the second question sheds light on what frictions are relevant for users. If international transfers and e-commerce are important use cases, reductions in exchange rate frictions might further increase adoption (recalling that in the theoretical model, exchange rate uncertainty reduces the utility of Bitcoin). If we do not see those use cases, we might conclude that the existing frictions are prohibitive, but this finding would also be consistent with a

scenario where consumers have alternative payment methods. Understanding the share of different use cases also helps predict the growth of transaction volume, which impact exchange rates in the theoretical model. The third question bears on what types of substitute payment technologies might compete with Bitcoin in the future (the outside option in the model).

4 Aggregate Analysis of Bitcoin Exchange Rates

In this section, we present evidence about the co-movement of exchange rates, transaction volume, and the effective supply of Bitcoin. The data used in this section were obtained from Blockchain.info, which makes the data available publicly on its website. We rely on Blockchain.info because this site provides an estimate of transaction volume excluding the major Bitcoin exchanges, and we believe their list of addresses associated with Bitcoin exchanges is more complete and accurate than ours. Blockchain.info also makes use of heuristics to remove change.

To operationalize the theoretical prediction that exchange rates are proportional to the ratio of volume to effective supply, we need to introduce the concept of Bitcoin velocity. In general, velocity is defined as the ratio of transaction volume to money supply. For the case of Bitcoin, we operationalize this as

$$\text{Velocity} = \frac{\text{Transaction Volume}}{\text{Exchange rate} \times \text{Supply of Bitcoins}}.$$

We can rearrange the velocity definition to solve for the exchange rate, as follows:

$$\text{Exchange rate} = \frac{\text{Transaction Volume}}{\text{Velocity} \times \text{Supply of Bitcoins}}. \tag{11}$$

This equation formalizes the idea that the exchange rate is the ratio of the demand (transaction volume) and the effective supply of Bitcoin. Note that there are no assumptions involved for this equation to hold; it merely rearranges an identity. Written this way, however, it is perhaps more natural to reason about the forces that affect the exchange rate: if the future supply of Bitcoins is known, then predicing prices boils down to predicting the ratio of transaction volume to velocity.

There are a variety of forces that can affect velocity of fiat currency. Digital currency has a distinct set of forces; for example, people may be more likely to hold their Bitcoins longer if they have few places to spend them or there are transactions costs involved in liquidating them. For Bitcoin, it seems likely that velocity and transaction volume would move together in the future: an increase in the ability to use Bitcoin would lead people to spend them more often, and also increase transaction volume. On the other hand, if transaction volume increases derive more from new adoption, then the transaction volume might increase substantially alongside more modest changes in velocity. Other factors that might affect the velocity of Bitcoin include: the share of users that are actively using Bitcoin (as opposed to users who may have received Bitcoin but aren't considering using them), the share of Bitcoins that have been lost (e.g. the passcode or wallet is lost) or confiscated, the opportunities to use Bitcoin commercially (e.g. merchant acceptance), and

the availability of convenient applications, e.g. on mobile platforms. We further observe that factors that might increase velocity could also affect transaction volume. In order to shed more light on these forces, in the next section we analyze transaction-level data.

Figure 3 shows the evolution of the market capitalization of Bitcoin (the product of the exchange rate and the supply of Bitcoins at that point in time, which is the denominator of (11)) alongside two measures of annual transaction volume in dollars (the numerator of (11)). The first is the total of the previous 365 days of volume, while the second is the total of the previous 30 days scaled up by 365/30. The Figure shows that market capitalization roughly tracks the volume of transactions. Figure 4 shows the implied velocity series for two versions of velocity, one constructed with the prior 30 days of transaction volume (rescaled by 365/30) and the other with the prior 365 days of transaction volume. The chart highlights that, outside of some periods of high volume, the implied velocity has been fairly stable. To interpret the velocity series, it is useful to recall that velocity is fairly poorly understood in the traditional economy. The M1 Velocity, which is equal to the ratio of nominal GDP and the M1 Money supply has varied between about 5 and 11 since 1970.[2] For Bitcoin, although we report measures that exclude transaction volume with exchanges as identified by Blockchain.info, we most likely do not have a measure of true economic activity comparable to GDP. In addition, the measurement issue is likely to overstate non-exchange transaction volume in times with a lot of speculative activity, because the default is that transaction volume is included in our measure. Since high Bitcoin prices have historically attracted a lot of investor activity, we might view the estimates with more skepticism during those times.

It is instructive to compare actual Bitcoin prices to those that would be predicted if Bitcoin velocity was held constant. The result of this exercise is shown in Figures 5 and 6. The charts show the prices predicted by 11 if velocity was held constant at a value of either 3 or 5. We see that outside a few periods of price spikes, the predicted prices roughly track the actual prices. In other words, the aggregate data is qualitatively consistent with the theory that prices are determined by current transaction volume and a fairly stable velocity. To quantify the goodness of fit, we conducted the following exercise. First, to avoid the problem that exchange rates in the recent past affect the measure of annual transaction volume, and exchange rates are serially correlated over time, we attempt to predict prices 30 days ahead. We consider measures of the total supply of Bitcoins, the 365-day transaction volume, and the ratio of the two as predictors. We further aggregate the data by taking weekly averages. Since exchange rates fluctuate greatly in levels, we analyze goodness of fit in terms of logarithms of these weekly averages. Specifically, in a regression of the logarithm of 30-day ahead exchange rates on the logarithm of the ratio of 365-day transaction volume to Bitcoin supply using data from 2013-2015, the R^2 is .83. To see how predictable the prices are just using time alone, we consider a regression of 30-day ahead exchange rates on week number (starting in 2013), as well as week number squared, cubed, and to the fourth power, as well as the logarithm of the week number. The sum of squared residuals from the latter regression is 3.5 times greater

[2]See https://research.stlouisfed.org/fred2/series/M1V. The M1 money supply includes physical cash, demand deposits, checking accounts, and other liquid accounts.

than in a regression that also includes the ratio of 365-day transaction volume to Bitcoin supply as a covariate. The R^2 from the latter regression is .97.

A better understanding of the microeconomics of Bitcoin use would shed more light on the sources of fluctuations and time trends in velocity and thus price formation. In addition, an examination of transaction-level data would help answer the question of whether the data are consistent with the theoretical model where transaction volume is the primary determinant of exchange rates; if speculation and investment play a more important role, then the theoretical model that incorporates speculators may be more appropriate. We proceed with this analysis in the next section.

5 Micro-analysis of the Public Ledger

5.1 Entity Detection and Analysis

The Bitcoin ledger is public, which might make it seem as though analyzing transactions would be fairly straightforward. A number of special features of the system make it more challenging to translate the ledger data into economically meaningful quantities, and indeed it would be impossible to do so with perfect accuracy. Thus, we will follow the literature and employ a number of heuristics to conduct our analysis. For this reason, it must be emphasized from the beginning that all of our empirical conclusions should be interpreted as noisy measures of the truth. In addition, applying different heuristics could change the conclusions.

The Bitcoin blockchain and its institutional features have been described by a number of authors; see Brito and Castillo [2013], Böhme et al. [2015]. Here we will outline a few key features that affect our analysis. First, the blockchain records transactions between distinct Bitcoin addresses. These addresses are typically managed through a software product called a "wallet," where a wallet contains multiple addresses. The addresses are pseudonymous–no individual identities are attached to them– and in addition, the addresses are not associated with their corresponding wallets in the Bitcoin ledger. It may be very difficult to figure out whether two wallets belong to the same person, and we will not attempt to do so in this paper. Instead, we use heuristics to assign different addresses to the same wallet.

As a basis for our heuristics for (probabilistically) grouping addresses into wallets, we make use of the fact that when someone makes a Bitcoin transaction, all of the Bitcoins from that address are used as an input to the transaction. Those funds are split into the portion that goes to the receiving address, and coins that are returned as change, where the protocol produces a new change address (Cha). If we can identify which part is change, we can associate the address that is the recipient of the change with the wallet of the sender. We also incorporate a few other heuristics. We refer to the group of addresses associated together as an "entity," where the heuristics are intended to align entities with wallets in the case of individual users (for firms, we may combine wallets from a single firm into one entity).

A brief overview of our heuristics is outlined as follows:

1. CoinJoin and Mixing Transactions: We use heuristics to identify "mixing" transactions ([Coi]).

If there are more than four inputs and more than four outputs, these are identified as mixing (discussed further below), and these transactions are excluded from other heuristics. In our data, 14,957,194 addresses are identified to be part of at least one mixing transaction in our data. 66,160,456 addresses are in at least one transaction not marked as mixing.

2. The Common-senders heuristic ([Ron and Shamir, 2013]): In a m-to-n transaction, we assume that all m input addresses are controlled by a common entity. Furthermore, if one input is in two separate m-to-n and m'-to-n' transactions, then we assume that all (m + m') inputs are controlled by the same entity.

3. Change identification: An address designated likely to be the change address is grouped in the same entity as the sending addresses.

   (a) In a two-output transaction, if one of the outputs has 3 more decimal places than other output value (which has 4 or fewer decimal places), we declare the output with the larger number of decimal digits to be the change address. For example, if a transaction has outputs .02 and .01615265, we declare the latter to be the change address. This is based on the assumption that Bitcoin users are unlikely to send amounts to other users in cognitively-difficult amounts with a high number of decimal digits.

   (b) (Meiklejohn et al. [2013]): If there is only one address that is first seen as an output in a non-generation transaction for the first time, and there are no self-change addresses in this transaction, we say this address is a one-time-change address.

Relative to Meiklejohn et al. [2013], we add the Mixing and decimal place heuristics. Our aggregates currently closely approximate the Estimated Transaction Volume on Blockchain.info (Blockchain.info). We also incorporate other sources of data about known entities. Meiklejohn et al. [2013] provided data about 1,138 addresses, including the major exchanges circa early 2013, Silk Road, gambling, and major vendors. This data was obtained by doing transactions. We used publicly posted lists of major known Bitcoin addresses; sources are detailed in Section 8.1.1 of the Online Appendix. Finally, we scraped the blogging website http://Bitcointalk.org, finding blogs supported by Bitcoin donations that indicated country of origin for the blog. In total, 3,906,194 known addresses were identified. Table 2 summarizes the number of addresses and the number of distinct names by industry, where our industry assignment was conducted manually using heuristics and internet research. Finally, we classify entities that are very large as unknown online businesses if we cannot identify them otherwise. We define unknown online businesses as the top 200 entities by transaction volume that both have transacted with at least 100 other entities and are not already known. Though this is by no means a comprehensive list of all of the businesses who use Bitcoin, we nevertheless believe this is a good heuristic based on Figure 10, which displays the number of entities with which each of the top 1000 known businesses by transaction volume with at least one counterpart have transacted. Around 75% of these known businesses have at least 100 counterparts, making it a reasonable assumption that entities with at least 100 counterparts and extremely large

transaction volumes are companies. Note that errors in merging addresses into entities will be carried into this definition of large unknown entities. See Table 1 for further details on the industries to which the businesses were assigned. We define "individual users" to be users that are neither known or unknown large entities, even though some of the individual users may be businesses of some type.

While the exchange of Bitcoins normally occurs on the blockchain, the public ledger where the transfer of Bitcoin ownership is recorded, there is some Bitcoin activity that can occur "off-chain," or not on the blockchain. For instance, a Bitcoin wallet service can transfer the ownership of the Bitcoins at a particular address from one wallet user to another. Though Bitcoins have effectively been given from one user to the other, there is no record of this transaction on the blockchain. Off-chain activity is nearly always facilitated by trust between the two parties exchanging Bitcoins, and when two parties do not trust one another activity usually involves a trusted third party (Off). Without the trust that debts will be repaid and that third-party services will return money users will not participate in off-chain activity. An example of an off-chain service is Circle, which allows customers to interchangeably send and receive traditional fiat currencies or Bitcoins after they have made a deposit of money. There are no fees to convert funds between currencies or to transfer money between users, and transactions are effectively debts transferred between users of the system–most transactions stay within the Circle system, with nothing written to the blockchain, and transactions are largely only made during deposits and withdrawals (Cawrey [2014]). We are not able to capture this activity in this paper.

5.2 Entity Creation: Summary of Initial Results

In Table 3, we summarize the results of our entity creation exercise. We show the number of addresses and entities, cateogorized by industry (based on the category of the known entity). We see that the average number of addresses per entity varies with industry type; for example, individual users have an average of 2.4 addresses per entity, with a median of 1. Since many addresses are involved in only one transaction (they receive Bitcoins, but never transact again), it is not surprising to see so many single-address entities (19,654,960 out of 27,474,538 entities).

Table 4 shows the fraction of transactions (as a share of the dollar value, and as a share of total transactions) taking place with each type of known entity. We see that transactions among individual users are the majority of transactions, and large unknown entities are the second largest category. Among transactions with known entities, exchanges are the largest category, followed by gambling, contraband, and mining. It should be noted that our ability to identify known entities is almost certainly not uniform across categories, so the statistics should be interpreted with caution.

5.3 Adoption and Usage

In this section, we consider adoption and usage patterns for different types of users. This has implications for the evolution of transaction volume, which feeds back into prices as well as volatility (which in turn affects usage).

We use heuristics to group users into types according to their behavior, such as whether they appear to be infrequent or frequent transactors, as well as their length of observed engagement with Bitcoin. Figure 7 shows how user types are defined.[3] Figure 8 shows the evolution over time of the fraction of Bitcoin balances associated with each user type. A key observation is that the short-term and long-term frequent transactor categories are not growing; which suggests that Bitcoin velocity may decline over time (potentially affecting future transaction volume as well as expectations of future transaction volume, and thus prices). Another observation is that the long-term infrequent transactors and investors together hold most of the Bitcoins. On the other hand, in terms of the share of entities, short-term users and one-time users together have transitioned from less than 20% of entities in July of 2012, to more than 40% of entities in May 2015.

We next consider how the fundamentals underlying velocity vary across user groups and over time. Since velocity is an aggregate concept, we create a definition that is easier to work with at the user level. We define "staleness" of Bitcoins at the level of the user. Using a first-in, first-out ordering, we measure the length of time each Bitcoin spends in the entity's wallet. In a given month, the average staleness of the Bitcoins for a group of users is defined as the average length of time that one of the users' Bitcoins has been in their wallet.

Figure 9 illustrates the staleness of Bitcoins over time. We see that the average staleness within the miner category has been growing steadily over time, presumably because a large share of mined Bitcoins remain in the wallets of the original miners and are not spent. For the remaining types, long-term infrequent transactors and investors have staleness growing over time, as well. Since many of these are newer to Bitcoin, the staleness of the investors is between one and two years as of May 2015, and investor staleness grows somewhat linearly.

Putting together Figures 8 and 9, we see that that much of the transaction volume is associated with one-time users. Over time the share of transaction volume belonging to miners has decreased steadily in terms of the outputs of transactions. Starting in mid-to-late 2014, the staleness of the Bitcoins held by the "other known entities" started significantly increasing, and while in May 2015 the average staleness of the group's Bitcoins was nearly 400 days, the average staleness had not passed even 200 days until November and December 2014. The decreasing activity levels of these "other known entities" can be at least partially explained by the fact that a significant part of our information on known addresses was provided by Meiklejohn et al. [2013], and in all likelihood many of the companies active in early 2013 were no longer active in 2015, leading to an increase in the staleness of the Bitcoins that they were still holding when they ceased operations. Finally, exchanges accounting for more than 10% of transaction volume in terms of both inputs and outputs as of May 2015. We break out the "old miners" from "new miners," where old miners are miners that created their accounts before the first halving day (November 28th, 2012), because some of the early Bitcoin miners may have lost access to their Bitcoins; the coins were not very valuable at the time and so it is plausible that some individuals may have been careless.

[3]The mixing user type has exactly two transactions, both of which are mixing transactions, so that the entity appears to be only used to move funds through the mixing service.

5.4 Financial Privacy and Illicity Activity

A next question about the use cases of Bitcoin concerns its privacy-preserving (and privacy-destroying) characteristics. Since there is a public record of all transactions, Bitcoin leaves a trail, unlike cash. On the other hand, since Bitcoin does not link user identity to addresses, it appears to preserve anonymity in other ways. One way that a user could be de-identified is if one of their associates sends them Bitcoin; and then the associate reveals their address in some other context. Law enforcement could potentially look for associates of a suspected criminal and attempt to learn their Bitcoin addresses through a variety of means. (For users of off-blockchain services such as Circle, users should presumably be aware that law enforcement might be able to elicit their personal information and transaction activity under a court order; and those services require identifying information for sign-up.)

Recognizing the potential privacy risks of leaving a permanent, public record of Bitcoin activity, Bitcoin industry participants have created a variety of services intended to protect the financial privacy of users. One example is a "mixing" service ([Coi]). In short, mixing services combine funds from a variety of Bitcoin addresses and then redistribute them to a set of new Bitcoin addresses. In principle, if 10 addreses put in 1 Bitcoin each and 10 addresses receive 1 Bitcoin each, it is difficult to ascertain which of the 10 addresses was the source for any recipient. Of course, the individuals must trust the mixing service to give them their Bitcoins back after they send them in; this is accomplished by reputation, and potentially by allowing the software that carries out the transaction to be inspected.

As described in Section 5.1, we use a heuristic to identify mixer transactions. In particular, if a transaction involves 4 or more inputs and 4 or more outputs, we label it a mixing transaction. An important point that affects the interpretation of all of our results is that once an entity has experienced a mixing transaction, we lose track of the funds that went into the mixing transaction. We may still follow the behavior of the original entity, but we can no longer connect post-mixing behavior for the coins that were a part of a mixing transaction to the original entity.

Figure 11 illustrates the adoption of mixing services over time. Adoption increased substantially after the wallet service Blockchain.info began offering a free mixing service in 2013, in terms of the share of transactions and transaction volume using mixing but especially in terms of the percentage of entities that had been the recipients of a mixing transaction.

Presumably, users who engage in illicit activity should be more likely to use mixing services. We find that contraband buyers are more likely (19% of all buyers) than the general population (12%) to use mixing to protect their privacy; gamblers, by contrast, are very similar to the general population, also having approximately 12% of all gamblers in Bitcoin history using mixing. In Figure 12, we show the difference between the percentage of all entities using Bitcoin and the percentages of "gamblers" and "contraband users" respectively. We break out the fraction of entities using CoinJoin by the number of distinct entities that the entity has transacted with (the "degree of connectedness" of the entity). It is possible for an entity to have a degree of connectedness of zero if all of its transactions are mixing (we cannot concretely identify distinct entities that an

entity is interacting with due to the very nature of mixing) or they do not make any transactions. Due to this fact, a very large percentage of entities with zero counterparts have engaged in mixing. Other than those with only one counterparty (where the users would not have much chance to do both mixing transactions and gamble, and are also not users that exclusively use mixing services for their transactions), gamblers have a higher share of mixer usage than other entities across the entire distribution of connectedness. For contraband users, the use of mixing services is even more pronounced compared to the average Bitcoin user; they display greater levels of mixing across all degrees of connectedness.

This evidence suggests that users are willing to take costly steps to protect their financial privacy when engaging in illicit activity in Bitcoin, and that illicit activity appears to be a use case that is consistent with user privacy, at least in the view of the users who do it.

5.5 The Transaction Network and Regional Classification

An interesting feature of Bitcoin is that, at least for transactions that occur on the public blockchain, it is possible to construct the network of Bitcoin entities, where we consider entities to be connected if they have ever engaged in a transaction. Our first finding (based on our entity definitions) is that because of the relatively low utilization rates of Bitcoin, the Bitcoin network is very sparsely connected. Less than 40% of entities have 3 or more connections, and only 10% have 7 or more. We now proceed to provide a more in-depth analysis of the nature of interactions among individuals in the network.

5.5.1 Community Detection

After constructing the network of Bitcoin entities, with edges weighted by the total transaction volume in US Dollars between two entities, community detection analyses can help us better understand patterns of which types of Bitcoin users interact frequently with each other. We use the Louvain network clustering algorithm (Blondel et al. [2008], using the software available at https://sourceforge.net/projects/louvain/) on the network of Bitcoin entities to help achieve this objective. This network clustering algorithm works by initializing each vertex of a graph in its own community or within manually seeded clusters, repeatedly moving each vertex to the neighboring community which results in the greatest increase in modularity until the network modularity reaches a local maximum, then treating each of the formed communities as an individual vertex, with weighted edges representing both edges within the community (a weighted self-edge) and edges between communities. This process of local modularity optimization is repeated on this newly formed graph. The alternating steps of the grouping of vertices within communities into a single vertex in a new graph and local modularity optimization are repeated until vertices can no longer be grouped together. It should be noted that this application is somewhat unusal, in that there is publicly available data about a network of interactions among tens of millions of individuals.

In this case, we included exchanges and individuals in the clustering, but omitted all other businesses. We seeded all vertices in individual clusters except for the 2858 entities for which we

already have a geographic label, a combination of individual users and exchanges. These entities were seeded together in four communities representing the four overarching regions. However, in the very first step of the algorithm many of these entities shift to other communities as the Louvain algorithm works to maximize the modularity of the cluster assignments. Figure 13 gives us a summary of the sizes of the clusters created by the algorithm, where 49% of entities are not placed in a cluster at all (and so are not represented in the figure), consistent with the fact that many entities do not transact. We see that the vast majority of clusters have very few entities, consistent with the sparsity of the graph. Thus, we do not see strong evidence that there are a large number of clusters of users using Bitcoin as a way to make payments among members of the group. We see that the clustering algorithm was effective at identifying groups who transact together, and that groups do concentrate their transactions inside the groups: for clusters of size 10 to 99 entities, 40% of entities have multiple transactions within the cluster, and 68% of the cluster's transaction volume is within-cluster. We also see that most of the small to medium sized clusters involve only individuals. So even though these clusters are relatively rare and acount for a small share of entities, they do seem to be realizing the use case of transacting among a set of individuals; most published applications relate to much smaller networks.

5.5.2 Regional Classification

Another question concerning use cases for Bitcoin concerns whether it is used for international payments. Although we do not observe identifying information about Bitcoin users, it is still possible to use a statistical model to classify entities into regions. To do this, we created a training dataset of Bitcoin addresses for which we knew the country of origin of the user. Our primary data source is a website called "Bitcointalk.org," where users often ask for Bitcoin donations to support their blogs. They may also indicate their country of origin. By scraping this website, we obtained a training dataset of 2995 addresses with country labels. We find the entities associated with these addresses in our data, and assign the country label from the website to these 2858 entities. We group the countries into nine sub-regions, and four regions, the Americas, Europe, Asia, or Eastern Latin America. We chose these four regions because the features that are most important for predicting region relate to the time of transactions and thus the time zone of the user. The mapping from countries to regions is given in the Online Appendix. Summary statistics for our training dataset are in Tables 5 and 6. Table 6 shows that our training dataset is not representative: it has many more long-term users and miners than the full dataset.

To train our country classifier, we used Breiman's random forest algorithm (Breiman [2001]). For each entity in the training dataset, the algorithm produces a vector of estimated probabilities that the entity is in each region. We begin by generating covariates (features) used for prediction. These features are generated from the transactions the entity made on the blockchain, and fall into a few categories: the time of day at which an entity makes transactions, the known entities with which an entity has transacted, the number and volume of transactions an entity has made, and the countries with which the entity is known to have transacted.

In addition, we can generate additional features using information about other entities in the same clusters as identified in our application of the Louvain clustering algorithm as described in Section 5.5.1. An example of a cluster-based feature generated in this manner is "Number of Transactions with European Entities for each cluster."

In training the random forest model, we used the implementation included in the R package RandomForest. We used balanced sampling from each of the 4 regions during bagging so that the underrepresented regions in the training data would have equal influence in the construction of the model; recall that we have no prior information about what the actual proportions of the regions are in our data, and our training data is not representative. After determining optimal tuning parameters for the model through parameter sweeps, we performed feature selection by first creating a model using all features available, then selecting the top 50 features by each of two metrics: the mean decrease in accuracy and Gini feature importance. The features selected for the model are included in the Online Appendix in Tables 14-16. The best model has 96 features: 40 entity-based features, and an additional 56 cluster-based features. Interestingly, the 40 entity-based features were nearly all time-based, but the cluster-based features also focused on the cluster's transactions with both known companies and with the 2858 entities in our labeled training dataset. For instance, the feature "Number of Transactions with European Entities for each cluster" is the average number of transactions for the entities within the cluster with one of the 1000 (European entities in the training dataset, and the feature "Indicator of transacting with MtGox for each cluster" shows the fraction of entities in the cluster that transacted with the exchange "MtGox."

Table 8 shows the share of entities predicted to be in each region and the shares from our training dataset (recalling that we used equal-sized subsamples from each region when training the model). We see that Asia and especially Eastern Latin America comprised a much greater share of the entities predicted across the entire dataset than in our training dataset, and the opposite was true with the Americas and Europe. We evaluated performance using Out-Of-Bag (OOB) error;[4] The OOB error for the model is shown in Table 7. The model's average error rate of 0.402 across the four regions is far better than the 0.75 error rate one would expect if classifying four equally-sized classes randomly. Indeed, it may be surprising that the error rate is as low as it is, given the small size of the training dataset and the fact that a large share of users engage in only a few transactions per user. Our ability to predict well with a simple model and a small, non-representative training dataset suggest that if governments or regulators sought to undertake a more aggressive attempt to learn about users, it would be possible to gain substantial insight.

However, the error rate from our classifier is large enough that we explicitly account for it when using our predictions for further analysis. More precisely, we interpret the probabilities assigned by the random forest as posterior probabilities from an empirical Bayesian model (for more discussion of this in the context of random forests, see Taddy et al. [2015]). In conducting our analysis, we sample 100 draws from this posterior for each entity, calculate the relevant sample statistics, and

[4]Each tree in the random forest uses only a subset of the observations to estimate the tree; to calculate the OOB, we make predictions for each observation using only trees in which that observation was not included in estimation, so that the prediction does not depend on the label for that particular observation.

then report means and standard deviations from those samples.

Another concern with our regional classification approach is that for some of our analyses, it may seem circular that we use cluster-based features. For example, in analyzing cross-region flows, we are interested in whether users from different regions transact with one another. The community detection algorithm will place entities into the same cluster if they transact with one another; and some features used in the prediction algorithm are defined at the cluster level, so that they will be common across all entities in a cluster. Thus, entities in the same cluster are more likely to be assigned to the same country than entities from different clusters. To address this issue, to address robustness of our findings, we also created an alternative "clean" predictive model, one that does not use any features based on clusters. The predictive accuracy of this model is worse than our "best" model: the OOB classification error rate is .551. The qualitative results from the clean model are similar, but the results are noisier; these results are included in the Online Appendix. One of the biggest systematic differences is that the clean model misclassifies the Americas much more often (error rate .61), and so the Americas are "mixed in" with other regions much more in this model. This leads Asia to stand out more in terms of adoption from the rest of the world.

5.6 Cross-Region Flows and Regional Differences in Behavior

5.6.1 Adoption and Usage over Time

There are a number of reasons to expect that Bitcoin adoption might be correlated within a geographical area. First, friends might inform one another about Bitcoin or teach one another how to use it. Second, some geographies might have higher concentrations of individuals who might value Bitcoin, either because they are technologically savvy or curious, or because they have needs that might be served by Bitcoin. Third, Bitcoin is much easier to use if there is a well-functioning, trusted Bitcoin exchange that accepts wire transfers or clearinghouse transfers from local banks. Thus motivated, we examine regional adoption curves. Figure 14 reflects the regional adoption curves, where we calculate the number of entities with non-zero balances in each region for each of the 100 simulation draws, and then report the averages. The figure shows that the Americas and Asia had significantly more entities than Europe and Eastern Latin America since the creation of Bitcoin, and in particular adoption tipped up earliest in the Americas. It appears that Asian usage of Bitcoin increased significantly relative to usage in the Americas as of late 2013. This matches the reported expansion of the digital currency market in China in late 2013 (Luedi [2016]). The random forest model identifies Asian entities through time of day features as well as features such as the "Indicator of transacting with BtcChina for each cluster," where BtcChina was the most popular exchange in China in late 2013 (Luedi [2016]). Due to the very large number of entities, almost all of the differences in the curves in Figure 14 are statisticallly significant, in the sense that one can reject the hypothesis that the differences in averages are due to misclassification error. Standard deviations for cross-region differences are illustrated in the Online Appendix.

In Figure 15, we illustrate the differences between regions (with the Americas as the base region) in terms of Bitcoin balances. The variation in these balances across simulation draws is much higher

than the variation in the number of entities, because Bitcoin balances are highly skewed, with a relatively small number of entities holding a large share of balances. For instance, the first 100,000 entities hold 92% of all bitcoins. While the total balance curves relative to the Americas for each region show a similar story to the adoption curves, only one conclusion is statiatically significant: by 2015, Asia held a greater balance of Bitcoins than both Europe and Eastern Latin America, consistent with the upwards trend of Asian Bitcoin adoption highlighted earlier. However, a look at staleness, the amount of time for which an entity has been holding the Bitcoins they currently possess, gives us greater insight: in Figure 16, we see users in the Americas have started holding their Bitcoins significantly longer than the other 3 regions since early 2014. While users in the Americas are still holding their Bitcoins, their actual usage has significantly declined relative to the rest of the world. This is generally consistent with the finding that most users are short-term users. It is possible that more American users are miners who hold onto their Bitcoins or investors (see 7) than the rest of the world; we return to this question below.

5.7 Cross-Regional Flows

A key question concerns whether individuals are using Bitcoin for international payments, since this is a use case that is underserved by existing financial institutions. Although our classification only identifies regions, so that we will understate international transfers, cross-regional transfers are clearly international. We analyze cross-regional flows by comparing actual regional flows to counterfactual regional flows, where in the counterfactual we assume that the share of entities in each region is the same as that predicted by the random forest, but where in the counterfactual world, entities are randomly assigned to regions (at the rates predicted by the model for the full sample, as shown in Table 8) rather than receiving the probabilities assigned to the individual entities by the random forest model. We compare this counterfactual to the actual percentage of transaction volume that occurs across regions. We calculate the latter quantity by considering our dataset of 100 draws of regional assignments for each entity. We take the mean (and standard deviation) over the 100 simulation draws of the share of transaction volume that involves each pair of regions. Table 9 shows the difference between the actual percentages and the counterfactual percentages. We see that transactions with the Americas are lower across the board than one would expect if entities were randomly assigned to regions. In addition, East Latin America transacts more with Europe and Asia (as well as within East Latin America) than in the counterfactual world, consistent with the use of Bitcoin for international payments from that region, where access to banking is lower than the rest of the world. Unfortunately, we were not able to accurately separate out South America from North America with our classifier, where remittances to South America is one plausible use case for Bitcoin.

5.7.1 Interaction with Industries

We can also consider how Bitcoin is being used in different regions. Table 10 displays the fraction of transaction volume between the Americas and a given industry[5] (relative to total transaction volume with industries), and it also displays the difference between the fraction of transaction volume belonging to a given industry for the Americas and for the other three regions. By far the most transaction volume occurs with exchanges, and the "unknown" industries are a clear second. Only a few of the regional differences are statistically significant; Asia and East Latin America appear to do less business with merchants (although it is possible we have a less comprehensive dataset of known merchants in Latin America), and Asia appears to have a lower share in gambling and contraband.

Another way to look at the interaction between individuals and companies is to look at the minimum network distance between each entity and a company belonging to a given industry. The Bitcoin network can be represented as a graph either with edges as transactions, and the network distance is the number of edges between two nodes, representing entities. As finding the network distance between every entity in the Bitcoin network and each of 13 industries is extremely computationally intensive, we instead sampled 10,000 entities from the Bitcoin network as representative of the entire network. We sampled these entities weighted by the total lifetime transaction volume of the entity in US Dollars. Table 11 shows the results. The table shows the raw distances from entities in the Americas, and also shows the differences in distance between the Americas and each of the other three regions. Not surprisingly, exchanges are the "closest" to a given entity in the network, showing that they are the industry with which Bitcoin users most frequently interact. Perhaps more surprisingly, the network distance between the Americas and each of the industries is significantly greater than for the other three regions. We also see that Asia is similar to the Americas in terms of illicit activities such as gambling and contraband, while the other two regions are significantly closer, although the magnitudes of the differences are not large.

5.7.2 User Types by Region

While adoption curves helped to reveal the number of users from each region who started using Bitcoin over time, looking at the breakdown of users in each region by user types (defined in Table 7) helps us understand how Bitcoin users differ in different regions of the world. Table 12 shows the fraction of users from the Americas that belong to each user type as well as the difference between the percentage of entities belonging to a given user type in the Americas and the percentage of entities belonging to that user type within another region. The region with the most mixing entities relative to the total entities in the region is the Americas. If we can assume either that entities transact more within their own region or that the algorithm assigned these mixing entities to the Americas because they mainly transact with other entities from the Americas, this potentially explains how the Americas had a significantly higher network distance to various industries than other regions–

[5]Table 1 further explains the various types of businesses identified in the Bitcoin network.

they may have been transacting with mixing entities first to help anonymize their transactions. Another difference of note is that generally the Americas have a slightly greater proportion of their entities defined as "other known entities," meaning that the pool of companies identified in this paper is slightly more concentrated in the Americas than the overall population of entities. Finally, there are more long-term frequent transactors in regions of the world other than the Americas, consistent with the idea that other parts of the world have more use cases for Bitcoin, perhaps due to greater need for international transfers or reduced access to banking.

6 Conclusion

In this paper, we have developed a theoretical framework for Bitcoin adoption and Bitcoin pricing, and then used that framework to guide an empirical analysis. Our theory suggests that Bitcoin prices may in principle be pinned down by economic fundamentals, and aggregate data about how Bitcoin prices have changed over time are consistent with this idea. A more nuanced look into the micro-data of individual transactions reveals that many Bitcoin users are not active, and that many buy and hold Bitcoins. We use a variety of empirical techniques to draw inferences about use cases of Bitcoin and how they vary with different regions of the world. We provide evidence that illegal activity is one use case, and that users attempt to protect their financial privacy when engaging in those activities. We also show that it is possible to learn about user characteristics such as their region of residence simply by examining their Bitcoin activity, particularly the time of day when they transact, and the entities with which their "neighbors" in the network transact. Using a simple classifier, we are able to draw conclusions about international flows of Bitcoin as well as cross-regional differences in its use.

Our paper relates to a variety of broad themes in the study of information technology adoption and usage. First, our theoretical model highlights the important role played by frictions in adoption, ranging from the risk of failure of a technology, to the fact that the value of the technology is affected by the adoption of other users (through the exchange rate). These themes are important for many new technologies, and the model and evidence highlight how they operate in the specific context of Bitcoin. Our micro-level empirical analysis also highlights the fact that aggregate adoption statistics about a new technology can be misleading: many new adopters do not appear to find the technology useful on an ongoing basis, other than for saving. We also see that adoption and usage trends are somewhat localized, despite the fact that a key feature of the Bitcoin technology is that it is purely digital and independent of Governments. Our paper also relates to the literature on privacy; the correlation between illegal activity and protecting financial privacy shows that at least a subset of users appear to value their privacy, and that behavior responds to economic costs and benefits. Overall, Bitcoin presents a unique opportunity to observe both the adoption and micro-level user-to-user transaction and interaction data in the context of a new information technology product, and in an environment where the usage data is publicly available. Future research might further explore the network structure of Bitcoin and how users interact in the network.

References

bicoinwiki-change. Web. URL https://en.bitcoin.it/wiki/Change.

bicoinwiki-coinjoin. Web. URL https://en.bitcoin.it/wiki/CoinJoin.

Off-chain transactions. Web. URL https://en.bitcoin.it/wiki/Off-Chain_Transactions.

Distributed ledger technology: Beyond blockchain. UK Government Chief Scientific Adviser, UK Government Office for Science, 2015.

Elli Androulaki and Ghassan O Karame. Hiding transaction amounts and balances in bitcoin. In *Trust and Trustworthy Computing*, pages 161–178. Springer, 2014.

Elli Androulaki, Ghassan O Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. Evaluating user privacy in bitcoin. In *Financial Cryptography and Data Security*, pages 34–51. Springer, 2013.

Moshe Babaioff, Shahar Dobzinski, Sigal Oren, and Aviv Zohar. On bitcoin and red balloons. In *Proceedings of the 13th ACM conference on electronic commerce*, pages 56–73. ACM, 2012.

Alex Biryukov, Dmitry Khovratovich, and Ivan Pustogarov. Deanonymisation of clients in bitcoin p2p network. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 15–29. ACM, 2014.

Blockchain.info. Estimated transaction volume. Web. URL https://blockchain.info/charts/estimated-transaction-volume.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008 (10):P10008, 2008.

Rainer Böhme, Nicolas Christin, Benjamin Edelman, and Tyler Moore. Bitcoin: Economics, technology, and governance. *The Journal of Economic Perspectives*, 29(2):213–238, 2015.

Jeremiah Bohr and Muhammad Bashir. Who uses bitcoin? an exploration of the bitcoin community. In *Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on*, pages 94–101. IEEE, 2014.

Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A Kroll, and Edward W Felten. Research perspectives and challenges for bitcoin and cryptocurrencies (extended version). Technical report, Cryptology ePrint Archive, Report 2015/452, 2015.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, October 2001.

Jerry Brito and Andrea Castillo. *Bitcoin: A primer for policymakers*. Mercatus Center at George Mason University, 2013.

Daniel Cawrey. Are off-blockchain transactions bad for bitcoin? Web, May 2014. URL http://www.coindesk.com/block-chain-transactions-bad-bitcoin/.

Nicolas Christin. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd international conference on World Wide Web*, pages 213–224. International World Wide Web Conferences Steering Committee, 2013.

Pavel Ciaian, Miroslava Rajcaniova, and d'Artis Kancs. The economics of bitcoin price formation. *Applied Economics*, pages 1–17, 2015.

Joan Antoni Donet Donet, Cristina Pérez-Sola, and Jordi Herrera-Joancomartí. The bitcoin p2p network. In *Financial Cryptography and Data Security*, pages 87–102. Springer, 2014.

Jonathan Donier and Jean-Philippe Bouchaud. Why do markets crash? bitcoin data offers unprecedented insights. *PloS one*, 10(10):e0139356, 2015.

Gerald P Dwyer. The economics of bitcoin and similar private digital currencies. *Journal of Financial Stability*, 17:81–91, 2015.

Meghan Elison. New world, new rules. *Ripple Insights*, 2016.

David S Evans. Economic aspects of bitcoin and other decentralized public-ledger currency platforms. *University of Chicago Coase-Sandor Institute for Law & Economics Research Paper*, (685), 2014.

Stanley Fischer. Friedman versus hayek on private money: Review essay. *Jonathan Swift: The Critical Heritage*, page 206, 1995.

Neil Gandal and Hanna Halaburda. Competition in the cryptocurrency market. 2014.

Joshua S Gans and Hanna Halaburda. Some economics of private digital currency. *Rotman School of Management Working Paper*, (2297296), 2013.

C Gourieroux and A Hencic. Noncausal autoregressive model in application to bitcoin/usd exchange rate. *Econometrics of Risk", Series: Studies in Computational Intelligence, Springer*, 2014.

Hanna Halaburda and Miklos Sarvary. *Beyond Bitcoin: The Economics of Digital Currencies*. Palgrave Macmillan, 2016.

Jermain Kaminski and Peter Gloor. Nowcasting the bitcoin market with twitter signals. *arXiv preprint arXiv:1406.7577*, 2014.

Ghassan O Karame, Elli Androulaki, and Srdjan Capkun. Double-spending fast payments in bitcoin. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 906–917. ACM, 2012.

Narayana R Kocherlakota. Money is memory. *journal of economic theory*, 81(2):232–251, 1998.

Dániel Kondor, Márton Pósfai, István Csabai, and Gábor Vattay. Do the rich get richer? an empirical analysis of the bitcoin transaction network. *PloS one*, 9(2):e86197, 2014.

Philip Koshy, Diana Koshy, and Patrick McDaniel. *An analysis of anonymity in bitcoin using p2p network traffic*. Springer, 2014.

Ladislav Kristoufek. What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis. *PloS one*, 10(4):e0123923, 2015.

Joshua A Kroll, Ian C Davey, and Edward W Felten. The economics of bitcoin mining, or bitcoin in the presence of adversaries. In *Proceedings of WEIS*, volume 2013. Citeseer, 2013.

Jeremy Luedi. How 3 asian countries are reacting to bitcoin. Web, January 2016. URL http://globalriskinsights.com/2016/01/how-3-asian-countries-are-reacting-to-bitcoin/.

William J Luther and Josiah Olson. Bitcoin is memory. *Journal of Prices & Markets*, 3(3):2015, 2013.

Ramon Marimon, Juan Pablo Nicolini, and Pedro Teles. Money is an experience good: Competition and trust in the private provision of money. *Journal of Monetary Economics*, 59(8):815–825, 2012.

Sarah Meiklejohn and Claudio Orlandi. Privacy-enhancing overlays in bitcoin. In *Financial Cryptography and Data Security*, pages 127–141. Springer, 2015.

Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140. ACM, 2013.

Tyler Moore and Nicolas Christin. Beware the middleman: Empirical analysis of bitcoin-exchange risk. In *Financial cryptography and data security*, pages 25–33. Springer, 2013.

Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.

Micha Ober, Stefan Katzenbeisser, and Kay Hamacher. Structure and anonymity of the bitcoin transaction graph. *Future internet*, 5(2):237–250, 2013.

Fergal Reid and Martin Harrigan. *An analysis of anonymity in the bitcoin system*. Springer, 2013.

Dorit Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph. In *Financial Cryptography and Data Security*, pages 6–24. Springer, 2013.

Dorit Ron and Adi Shamir. How did dread pirate roberts acquire and protect his bitcoin wealth? In *Financial Cryptography and Data Security*, pages 3–15. Springer, 2014.

Peter Surda. Economics of bitcoin is bitcoin an alternative to fiat currencies and gold. 2012.

Matthew Taddy, Chun-sheng Chen, Jun Yu, and Mitch Wyle. Bayesian and empirical bayesian forests. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 967–976, 2015.

Lawrence J Trautman. Virtual currencies; bitcoin & what now after liberty reserve, silk road, and mt. gox? *Richmond Journal of Law and Technology*, 20(4), 2014.

Marshall Van Alstyne. Why bitcoin has value. *Communications of the ACM*, 57(5):30–32, 2014.

Aaron Yelowitz and Matthew Wilson. Characteristics of bitcoin users: an analysis of google search data. *Applied Economics Letters*, 22(13):1030–1036, 2015.

7 Figures and Tables

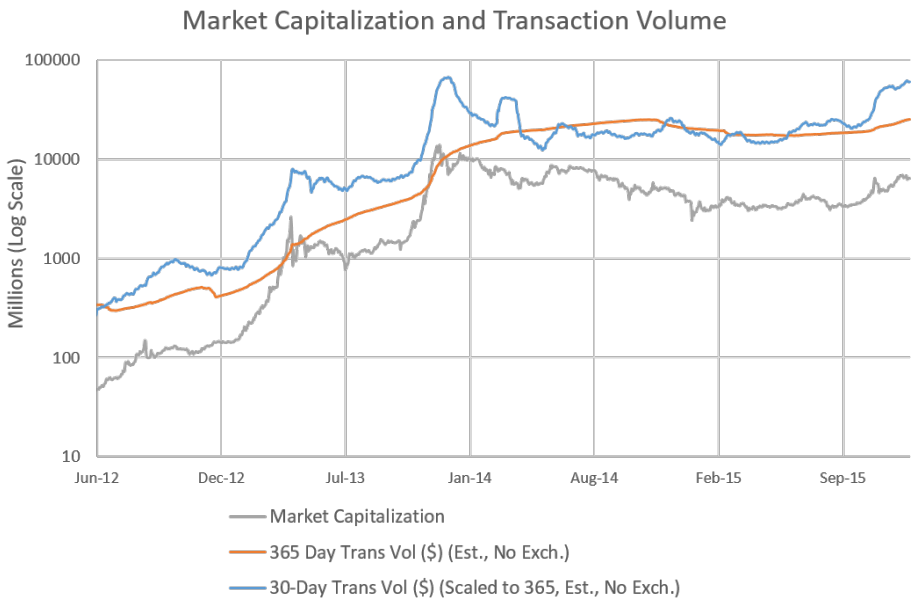


Figure 3: Market Capitalization is the product of the USD/Bitcoin exchange rate and the supply of Bitcoins. USD Transaction Volume excludes change using some proprietary heuristics as well as volume to and from the largest exchanges. We rescale 30-day transaction volume by multiplying by 365/30 so that it can be compared to aggregate transaction volume from the previous 365 days. All data from Blockchain.info.

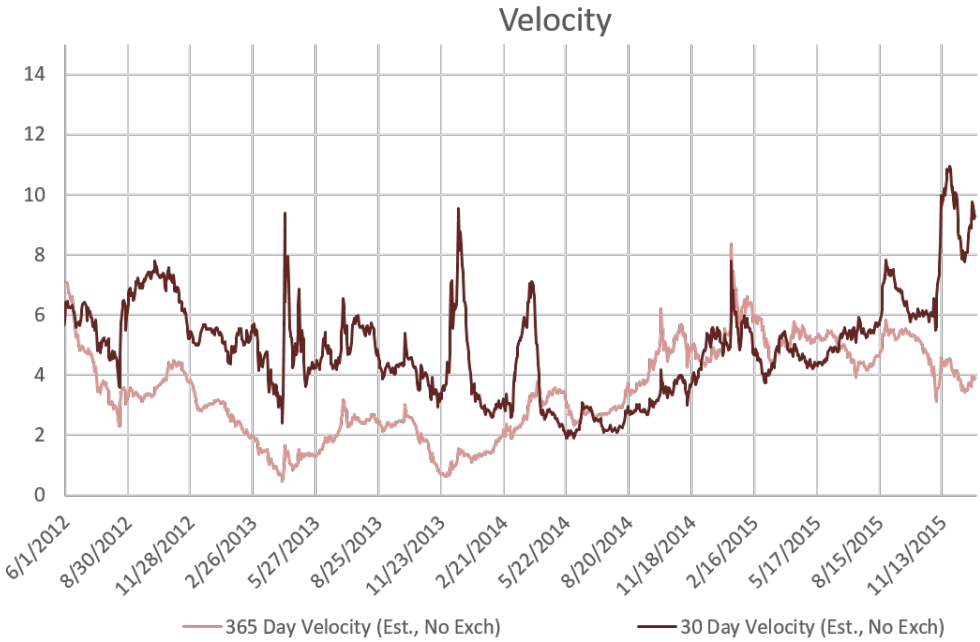


Figure 4: Velocity is the ratio between the transaction volume of Bitcoin and the supply of Bitcoins. Transaction Volume excludes change using some proprietary heuristics as well as volume to and from the largest exchanges. We show two series, one constructed with the prior 30 days of volume, and one with the prior 365 days of volume. We rescale 30-day transaction volume by multiplying by 365/30. All data from Blockchain.info.

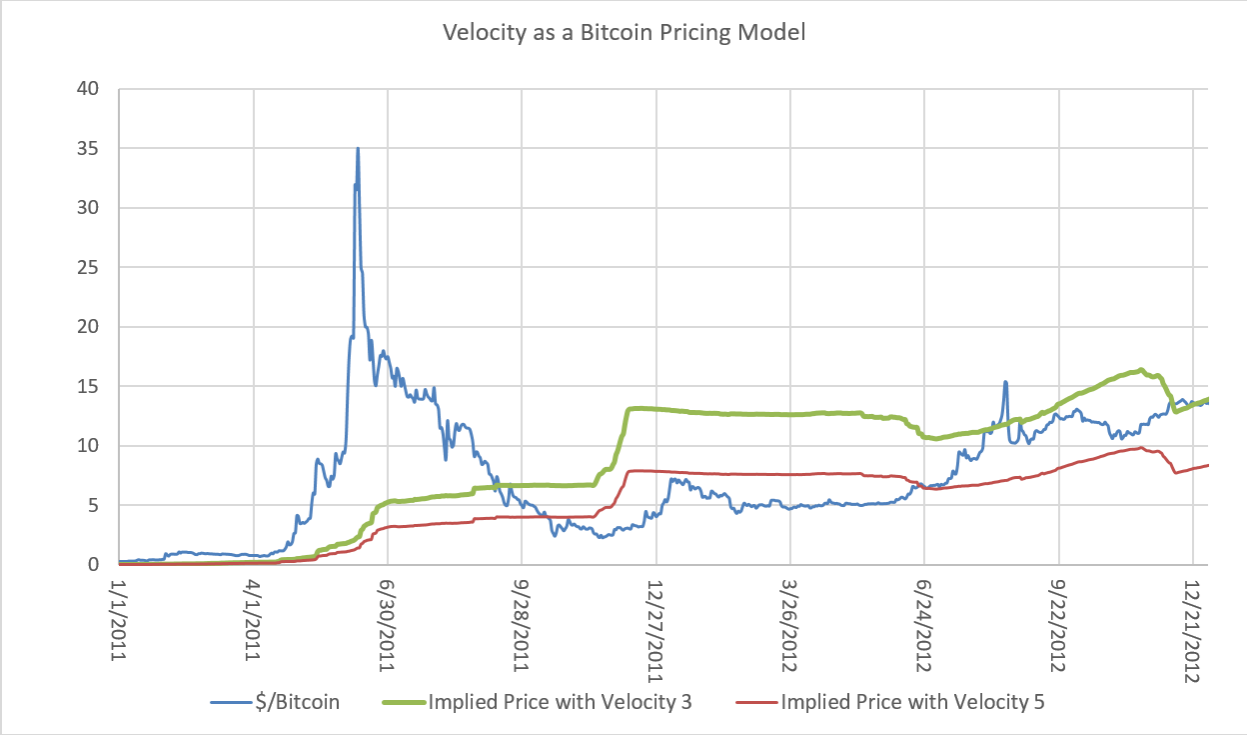


Figure 5: Implied Prices from Pricing Model versus Actual Prices, 2011-2012.

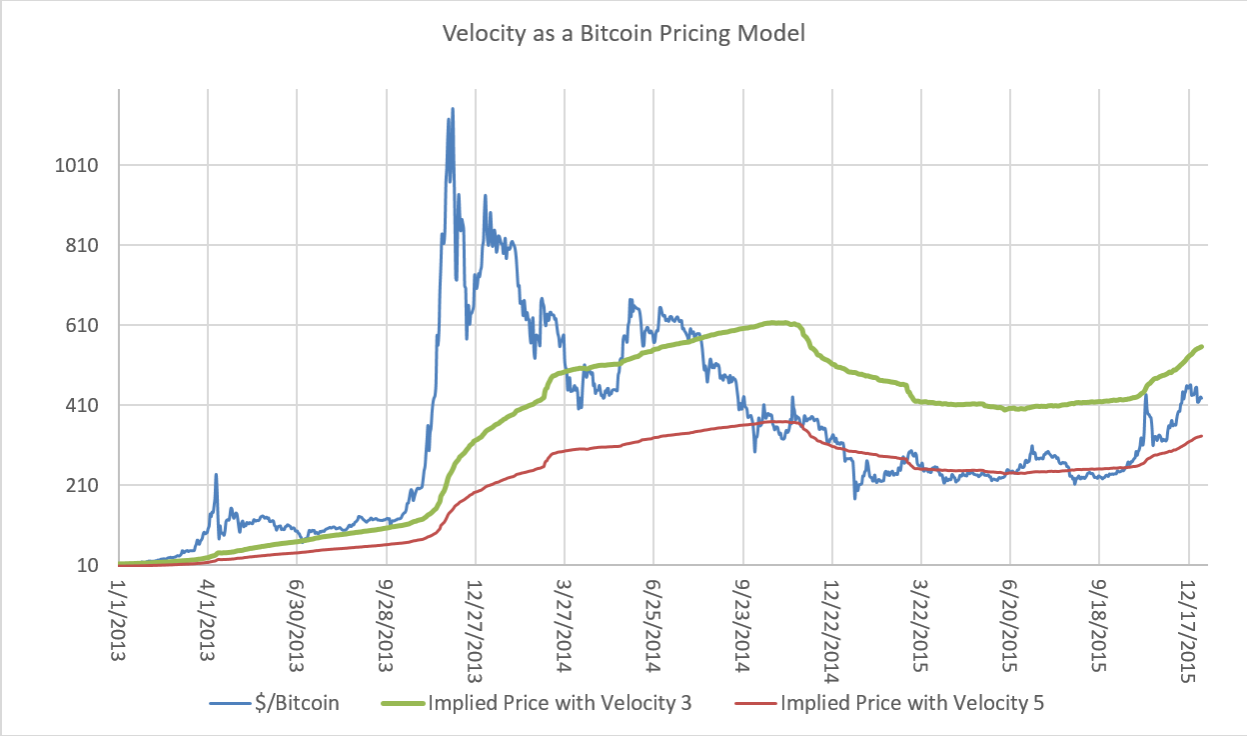
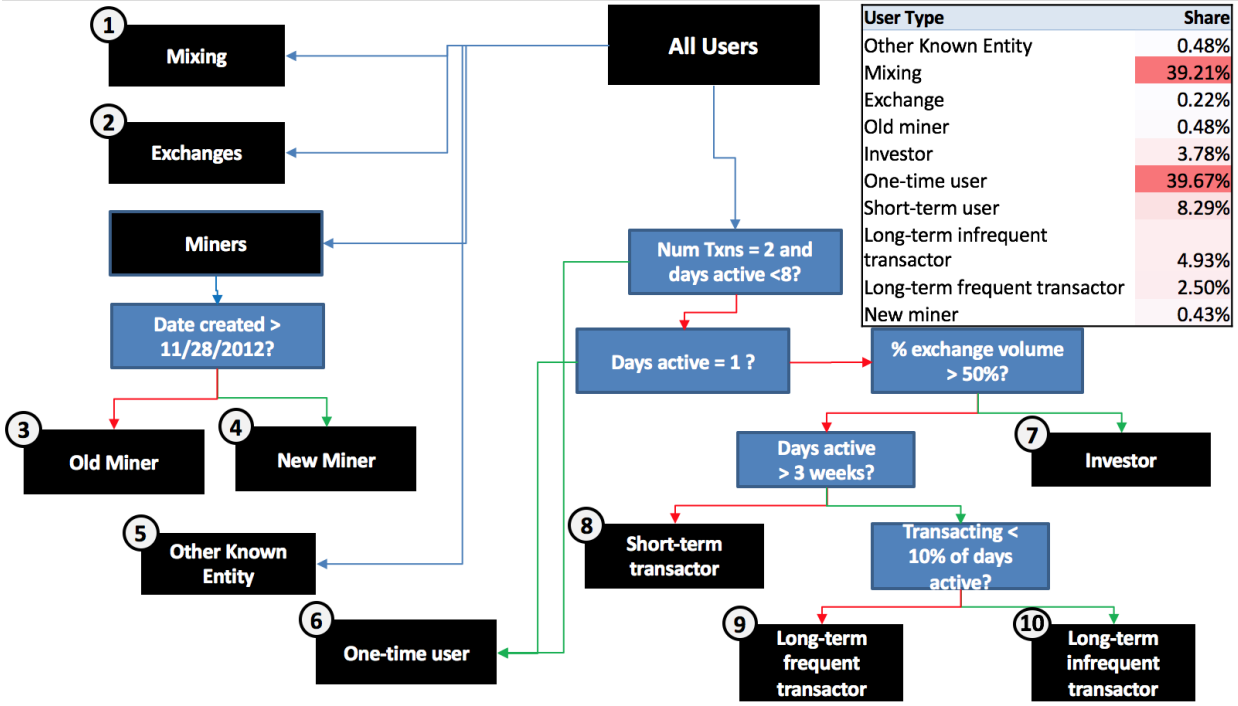


Figure 6: Implied Prices from Pricing Model versus Actual Prices, 2012-2015.



User Type	Share
Other Known Entity	0.48%
Mixing	39.21%
Exchange	0.22%
Old miner	0.48%
Investor	3.78%
One-time user	39.67%
Short-term user	8.29%
Long-term infrequent transactor	4.93%
Long-term frequent transactor	2.50%
New miner	0.43%

Figure 7: Definition of User Types.

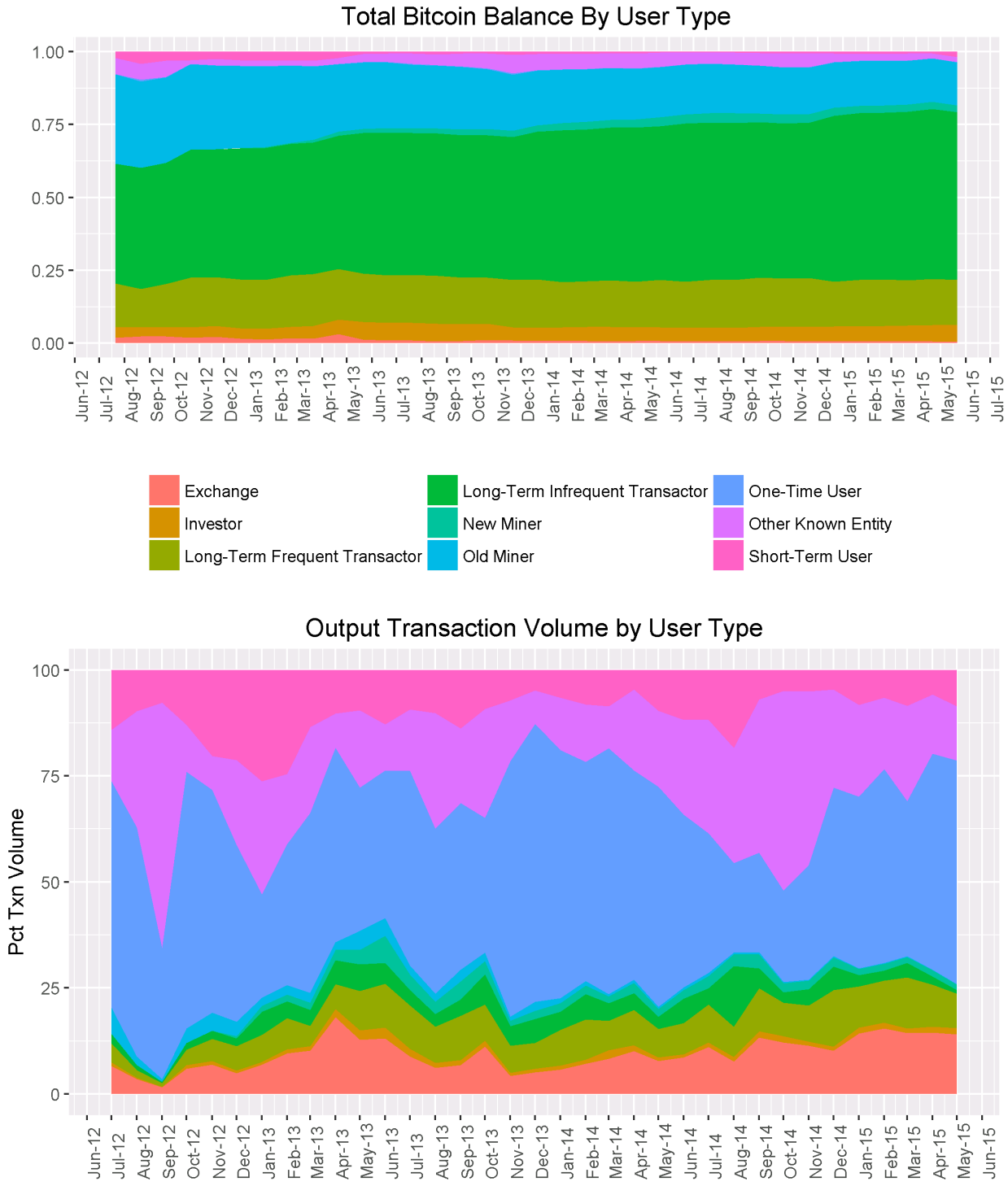


Figure 8: Bitcoin balances by user type (excluding mixing users), over time; Percentage of transaction volume by user type (excluding mixing users), over time, grouping by output entity. User types are defined in Figure 7.

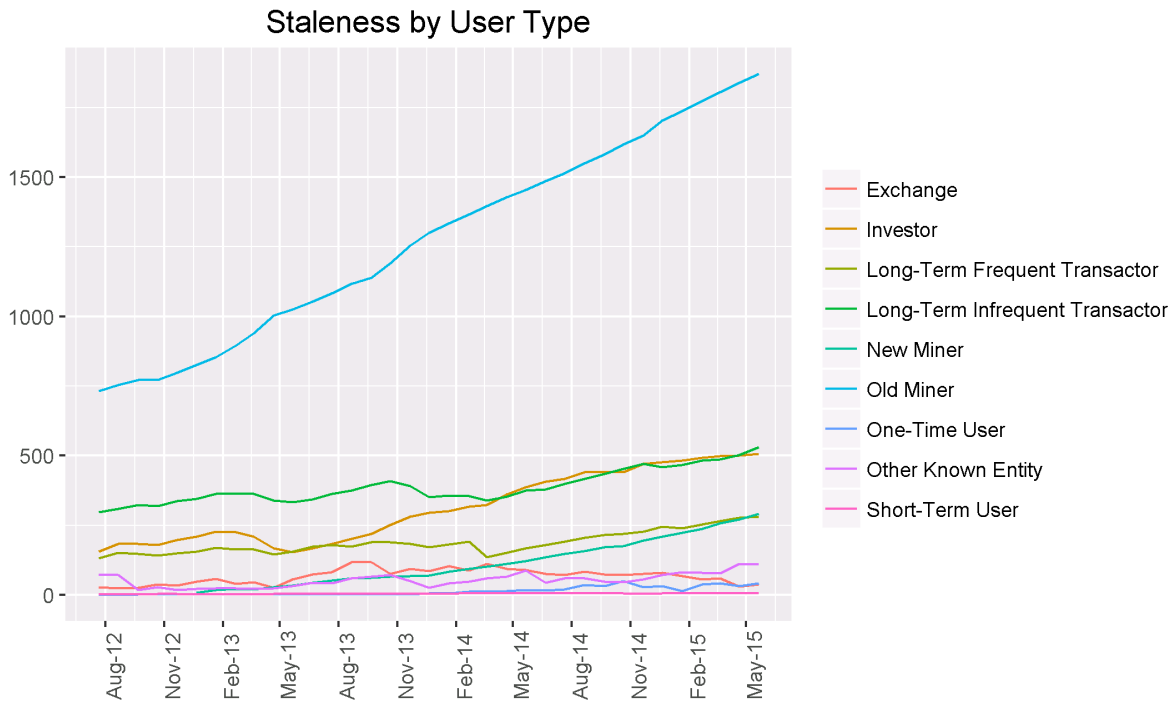


Figure 9: Staleness of Bitcoin by user type (excluding mixing users), over time. User types are defined in Figure 7.

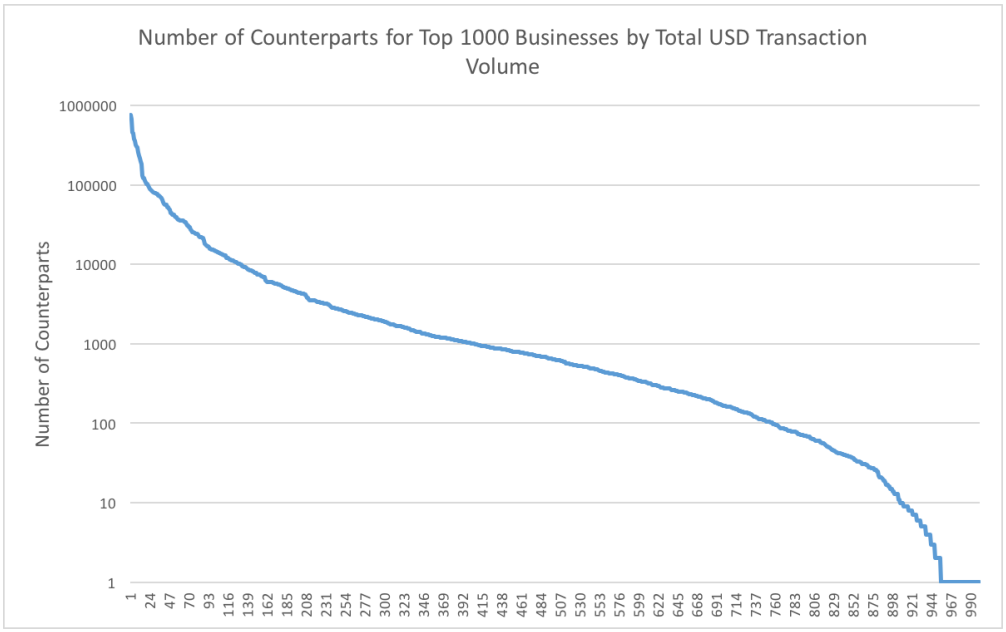


Figure 10: Number of counterparts for each of the top 100 known businesses with at least one counterpart.

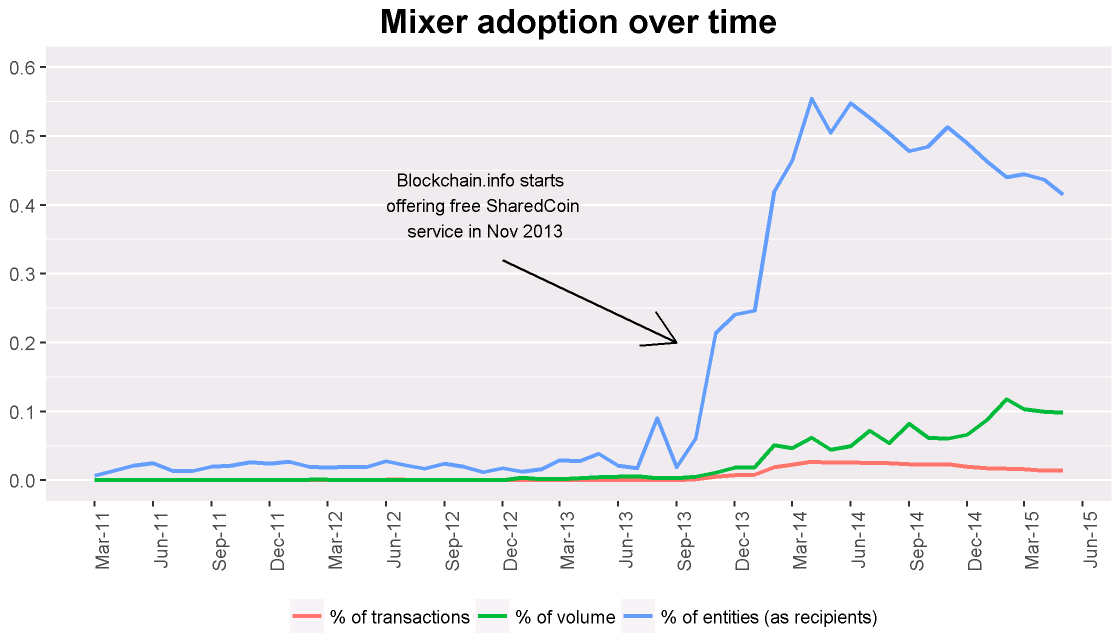


Figure 11: Estimated fraction of transactions and volume conducted by entities that have engaged in a mixer transaction.

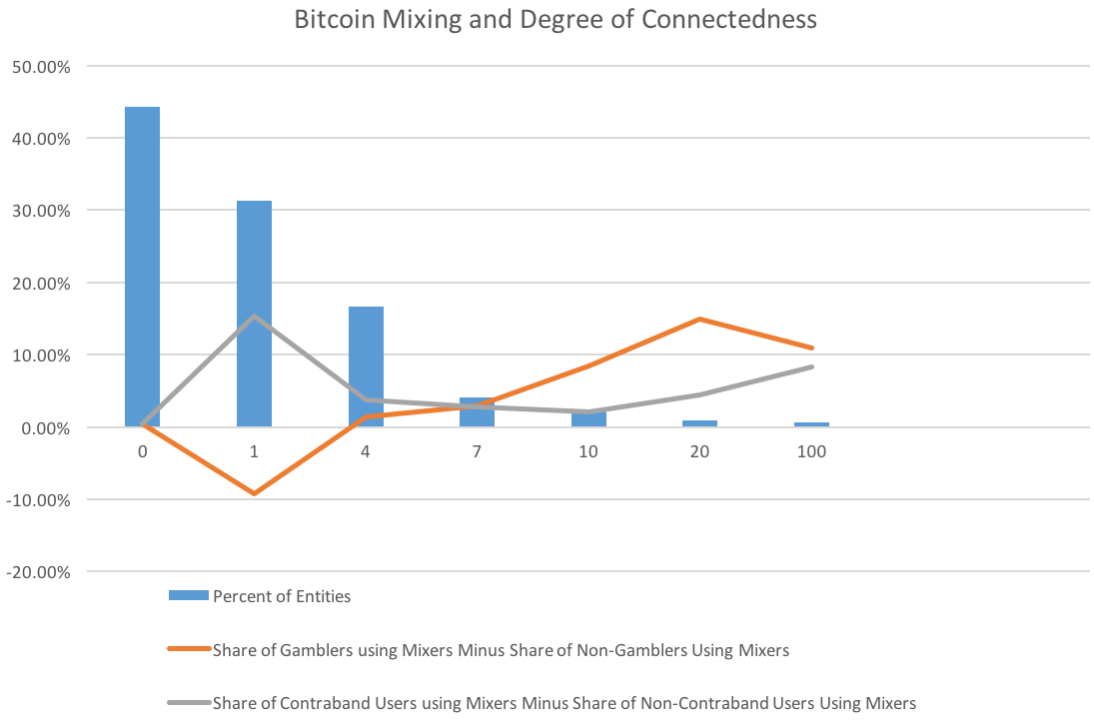


Figure 12: The difference between the share of gamblers using mixing services and the share of non-gamblers using mixing services, broken out by the number of other entities with which they transacted as well as whether they gambled or were involved with contraband goods. The analogous quantities are also presented for contraband users.

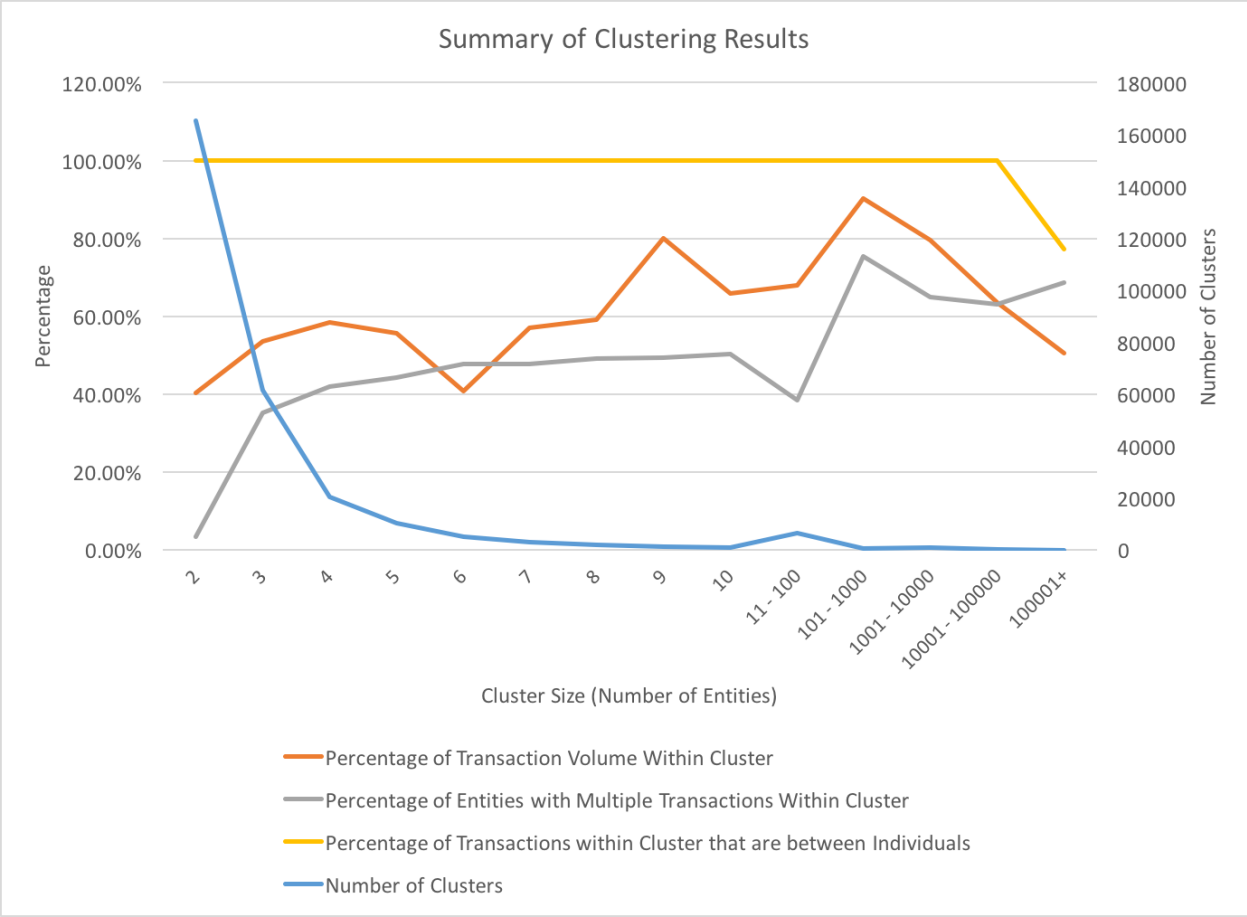


Figure 13: Histogram of cluster size and a number of metrics: number of clusters, percentage of transaction volume within cluster, percentage of entities with multiple transactions within cluster, and percentage of transactions within cluster that do not involve companies.

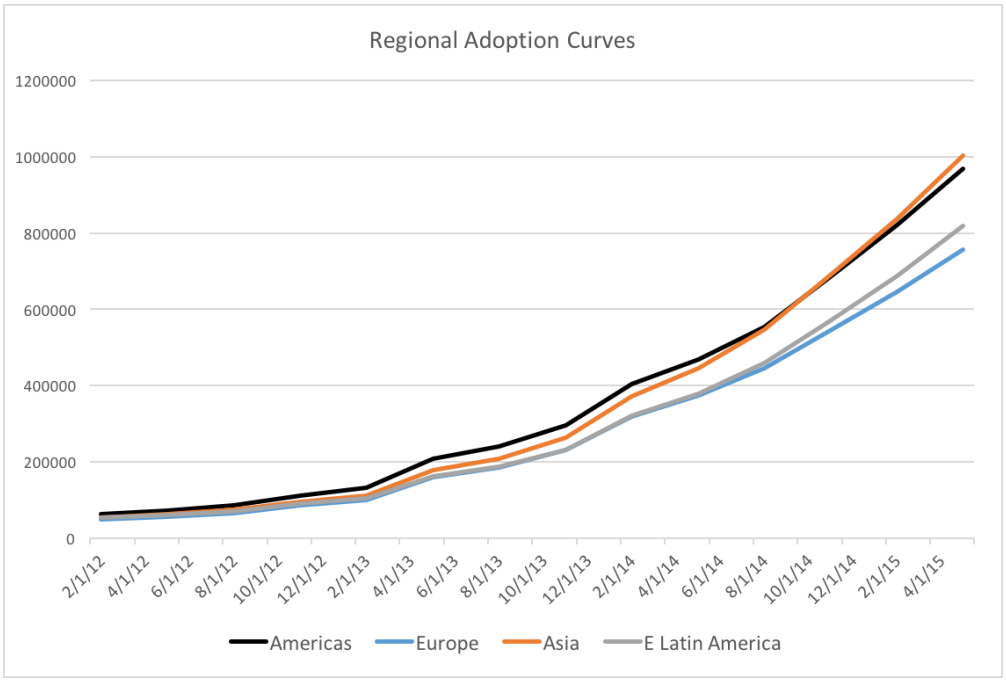


Figure 14: Number of entities in each region (from random forest assignments) with nonzero balance over time.

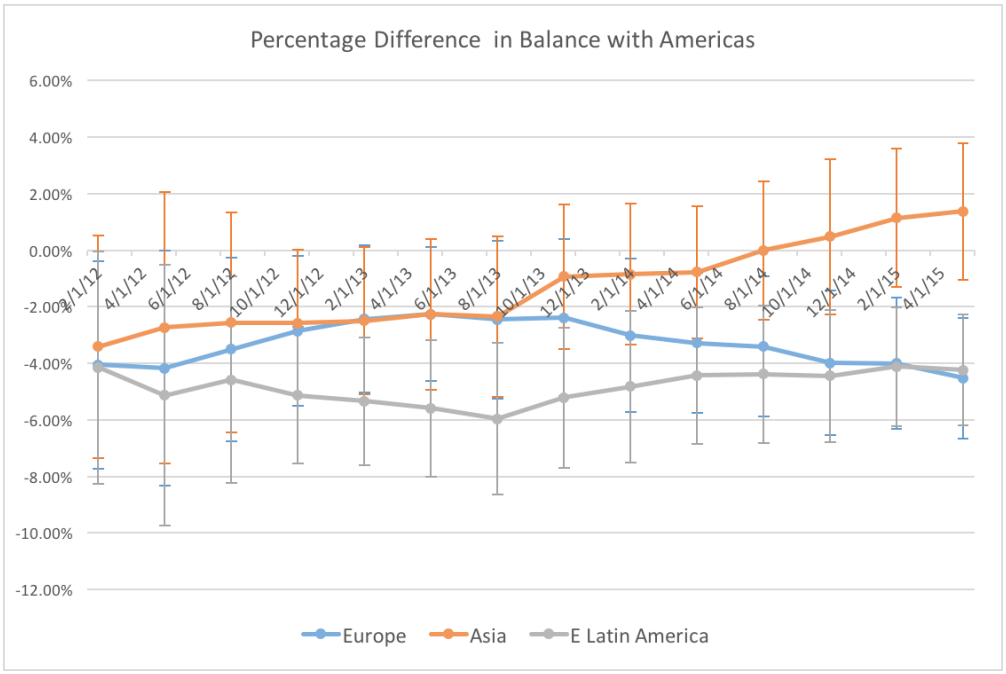


Figure 15: Percentage difference between total balance of all the entities in each region (assigned by the random forest model) and the Americas ((Balance in Region - Balance in Americas)/Balance in Americas)*100%.

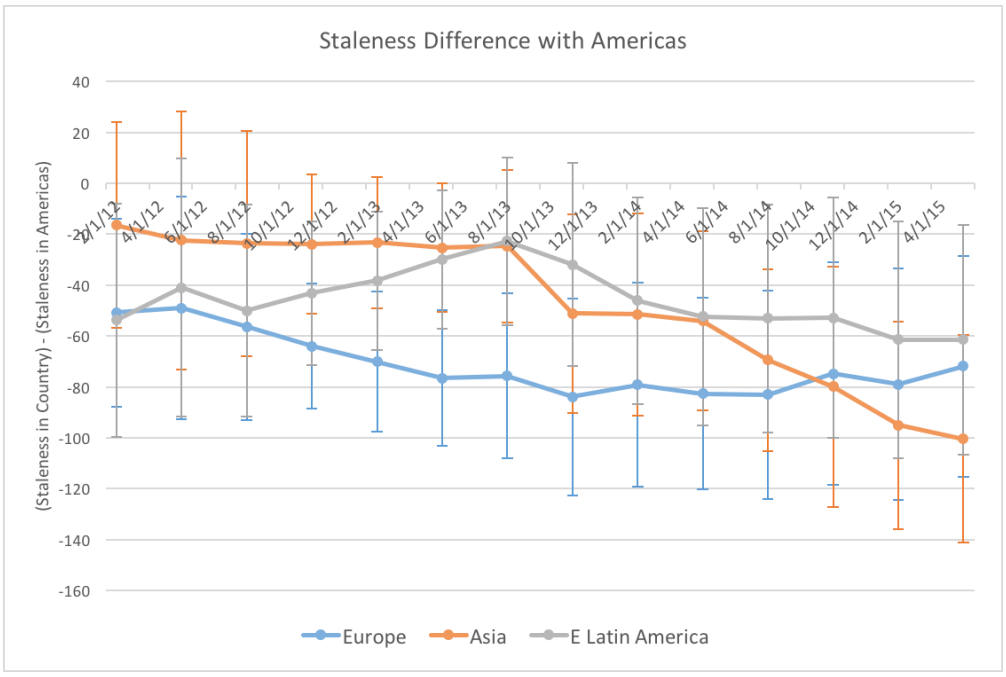


Figure 16: Difference between average staleness of Bitcoins held by entities in each region (assigned by the random forest model) and the Americas (Staleness in Country - Staleness in Americas).

Table 1: Types of Businesses Identified

Industry	Description
Gambling	Gambling services ranging from dice games to sports betting (ex. SatoshiDice, BetCoin)
Exchange	Services exchanging fiat currency for Bitcoins (ex. Mt. Gox, Huobi)
Contraband	Services selling illegal goods and drugs (ex. Silk Road, AgoraMarket)
Mining	Mining pool services that split mining rewards across its participants (ex. F2Pool, Bitminter)
Payments	Services facilitating Bitcoin use for businesses (ex. BitPay, CoinBox)
Wallet	Services helping individuals manage their Bitcoin addresses (ex. Instawallet, Flexcoin)
Laundry	Services that help obscure the exchange of Bitcoins between users (ex. Bitfog, Bitmix)
Giveaway	Websites giving visitors Bitcoins as a reward for completing small tasks (ex. BtcFaucet, Btcvisitor)
Ad Network	Web advertising services that accept Bitcoins as payment (ex. Coinad, Bitcoin Advertisers)
Merchant	Services that accept Bitcoins and sell a variety of goods (ex. Btcbuy, Bit Elfin)
Investing	Bitcoin lending and investing services (ex. Bitcoinia, Bitbond)
Ransomware	Malware that encrypts files on computers and offers to decrypt them in exchange for a Bitcoin payout, or ransom (ex. CoinVault, CryptoLocker)
Unknown	Top 200 entities by transaction volume that also transacted with 100 other entities and are not in the other categories

Table 2: Known Addresses by Industry

Industry	**# Addresses**	**# Distinct Names**	**Examples**
ad network	1804	3	CoinURL.com
contraband	901539	20	Silk Road, Evolution Market
exchange	1309859	144	Bitstamp
gambling	852239	151	Satoshi Dice
giveaway	11	3	Btc Faucet
Investing/ loans	51800	13	BitLendingClub
laundry	153642	7	BitcoinFog, BitLaunder.com
merchant	16534	18	BitMit, CryptoSexToys.com
mining	211752	32	Bitminter
payments	202102	13	BitPay
ransomware	1230	2	CryptoLocker
wallet	204631	19	Xapo.com

Table 3: Entities by Industry

	Addresses	Entities	Average	Median	AvgDailyTxns
Users	65,337,885	27,253,591	2.4	1	2.7
Known Companies	10,699,735	220,747	48.5	1	2.5
ad network	9,031	965	9.4	1	1.6
contraband	1,937,359	99,672	19.4	1	1.8
exchange	5,006,017	76,671	65.3	1	2.8
gambling	1,891,029	14,625	129.3	1	4.9
giveaway	54	5	10.8	8	20.4
investing	82,948	93	891.9	1	9.7
laundry	213,375	27,125	7.9	1	2.3
merchant	21,818	44	495.9	1	3.1
mining	956,378	1,054	907.4	2	11.7
payments	242,357	378	641.2	1	4.5
ransomware	1,367	2	683.5	683.5	6.4
wallet	338,002	113	2991.2	1	15.8
Unknown Companies	1,494,394	200	7472.0	575.5	296.7
Total	77,532,014	27,474,538	2.8	1	2.7

Table 4: Activity of Individual Users on Bitcoin

	Percentage of USD Transactions	Percentage of Transactions
Individuals/small unclassified	74.80%	55.76%
ad network	0.00%	0.80%
contraband	1.37%	3.08%
exchange	12.47%	13.94%
Unknown large entities	7.51%	1.75%
gambling	2.02%	17.11%
giveaway	0.00%	0.89%
investing	0.05%	0.11%
laundry	0.14%	0.24%
merchant	0.00%	0.03%
mining	0.83%	4.62%
payments	0.42%	0.72%
ransomware	0.00%	0.00%
wallet	0.38%	0.96%

	Sub-Region	# Addresses
1	North America	1312
2	Europe	1091
3	APAC	332
4	Australia and NZ	149
5	Former USSR	146
6	E.LatAm	133
7	South Asia	98
8	MEA	96
9	Latin America	50

Table 5: Number of Addresses by Region in Training Dataset

User Type	Share by Number of Entities, Full Dataset	Share by Balance, Full Dataset	Share by Number of Entities, Training	Share by Balance, Training
Exchange	0.36%	0.51%	1.82%	45.77%
Investor	6.23%	5.85%	11.21%	0.17%
Long-term Frequent Transactor	4.07%	15.37%	33.54%	0.75%
Long-term Infrequent Transactor	8.13%	57.13%	16.72%	1.7%
New Miner	0.7%	2.4%	17.1%	2.26%
Old Miner	0.79%	14.64%	7.12%	11%
One-time User	65.27%	0.48%	4.02%	0%
Other Known Entity	0.79%	1.51%	1.62%	38.35%
Short-term User	13.65%	2.1%	6.85%	0%

Table 6: Percentage of Entities of each User Type in Training Dataset, excluding Mixing Entities

	Predicted				
Actual	Americas	Europe	Asia	E. Latin America	Class Error
Americas	67.630%	13.594%	14.019%	4.758%	0.324
Europe	21.000%	52.500%	20.400%	6.100%	0.475
Asia	15.862%	13.621%	62.586%	7.931%	0.374
E.Latin America	15.842%	11.881%	16.832%	55.446%	0.446

Table 7: Random Forest OOB Accuracy

Table 8: Distribution of Regions Predicted by Random Forest Model

	Americas	Europe	Asia	E Latin America
Training Dataset	41.183%	34.990%	20.294%	3.534%
Predicted for Full Dataset	27.327%	21.321%	28.272%	23.079%
Predicted - Training	-13.855%	-13.669%	7.979%	19.545%

We compare the percentage of entities in each region in the training dataset with the percentage of entities assigned to each region by the random forest algorithm across the entire dataset, recording the difference between the two for each region.

Table 9: Cross-Regional Flow

	Region 2			
Region 1	Americas	Europe	Asia	E Latin America
Americas	-3.46651%*	-2.68761%*	-4.27628%*	-1.2179%
	(0.44624%)	(0.57293%)	(0.77542%)	(0.6837%)
Europe		0.95052%*	0.56011%	2.95052%*
		(0.46793%)	(0.77258%)	(0.87102%)
Asia			1.65714%	2.79778%*
			(0.88478%)	(0.76115%)
E Latin America				2.73223%*
				(0.81208%)

We first calculate the (counterfactual) share of transaction volume that would occur across regions if the region labels were assigned randomly to entities in proportion to the assignment proportions predicted by the random forest algorithm. We then calculate the actual share of transaction volume according to the assignments from 100 draws from the posterior distribution from the random forest algorithm. We report the difference between the mean of those draws and the counterfactual mean from random label assignment. We also report the standard deviation from the 100 draws.

* The asterisk denotes a result that is statistically significant at the 95% level.

Table 10: Industry-Region Flow

Industry	Americas	Europe Difference	Asia Difference	E Latin America Difference
Unknown	0.33852	-0.00451	0.05587	0.05235
	(0.07661)	(0.12046)	(0.10449)	(0.12226)
Gambling	0.08663	-0.02069	-0.03772*	-0.0263
	(0.01407)	(0.02003)	(0.01859)	(0.02174)
Exchange	0.45857	0.03071	0.01744	-0.00791
	(0.05586)	(0.08762)	(0.08492)	(0.07873)
Contraband	0.04421	-0.0033	-0.01473*	-0.00788
	(0.00286)	(0.0044)	(0.00384)	(0.00476)
Mining	0.03175	-0.00022	-0.00666	-0.00736
	(0.00527)	(0.00861)	(0.00676)	(0.00737)
Payments	0.01669	-0.00102	-0.0059	-0.00278
	(0.00516)	(0.00757)	(0.00687)	(0.00811)
Wallet	0.01324	0.00108	-0.00382	0.00027
	(0.00599)	(0.01069)	(0.00834)	(0.00795)
Laundry	0.00855	-0.00199	-0.00387	-0.00023
	(0.00265)	(0.00373)	(0.00311)	(0.00438)
Giveaway	0	0	0	0
	(0)	(0)	(0)	(0)
Ad Network	0.00009	0.00001	-0.00002	0.00001
	(0.00001)	(0.00002)	(0.00002)	(0.00002)
Merchant	0.00015	-0.00002	-0.00006*	-0.00004*
	(0.00001)	(0.00002)	(0.00002)	(0.00002)
Investing	0.00158	-0.00004	-0.00052	-0.00012
	(0.0002)	(0.00035)	(0.00029)	(0.00027)
Ransomware	0.00003	0	-0.00001	0
	(0.00002)	(0.00003)	(0.00002)	(0.00003)

We record the percentage of bitcoin flow involving the Americas that is with each industry. Then, for each other region we record the difference between the percentage of flow with a given industry for the given region and for the Americas

* The asterisk denotes a result that is statistically significant at the 95% level.

Table 11: Network Distance from Entity to Industry
Weighted by Entity Transaction Volume

Industry	Americas	Europe Difference	Asia Difference	E Latin America Difference
Unknown	1.944	-0.073*	-0.119*	-0.079*
	(0.015)	(0.023)	(0.022)	(0.026)
Gambling	1.98	-0.049*	-0.03	-0.072*
	(0.015)	(0.022)	(0.023)	(0.026)
Exchange	1.491	-0.088*	-0.076*	-0.066*
	(0.015)	(0.022)	(0.023)	(0.024)
Contraband	1.941	-0.069*	-0.013	-0.063*
	(0.017)	(0.025)	(0.025)	(0.029)
Mining	2.101	-0.074*	-0.048*	-0.088*
	(0.014)	(0.022)	(0.021)	(0.025)
Payments	2.121	-0.075*	-0.036	-0.084*
	(0.015)	(0.021)	(0.023)	(0.025)
Wallet	2.148	-0.07*	-0.041*	-0.088*
	(0.014)	(0.021)	(0.021)	(0.026)
Laundry	2.223	-0.077*	-0.047*	-0.085*
	(0.014)	(0.021)	(0.022)	(0.024)
Giveaway	2.398	-0.08*	-0.054*	-0.075*
	(0.015)	(0.022)	(0.022)	(0.025)
Ad Network	2.27	-0.073*	-0.052*	-0.083*
	(0.014)	(0.022)	(0.021)	(0.024)
Merchant	2.329	-0.075*	-0.044*	-0.072*
	(0.015)	(0.022)	(0.021)	(0.025)
Investing	2.261	-0.078*	-0.057*	-0.096*
	(0.014)	(0.021)	(0.02)	(0.025)
Ransomware	2.451	-0.078*	-0.025	-0.058*
	(0.015)	(0.023)	(0.023)	(0.025)

We measure network distance between the Americas and each Industry, then record the raw difference between the average distance to a given industry from a given region and the distance to the industry from the Americas, weighting averages by their total transaction volume with other entities

* The asterisk denotes a result that is statistically significant at the 95% level.

Table 12: Fraction of Entities of each User Type by Region

UserType	Americas	Europe Difference	Asia Difference	E Latin America Difference
Mixing	0.44855	-0.07106*	-0.07628*	-0.09686*
	(0.00018)	(0.00033)	(0.00031)	(0.00033)
Exchange	0.00236	-0.00017*	-0.00018*	-0.00032*
	(0.00001)	(0.00002)	(0.00002)	(0.00002)
Investor	0.03248	0.00693*	0.00739*	0.00874*
	(0.00005)	(0.00009)	(0.0001)	(0.00009)
Long-term frequent transactor	0.01934	0.00698*	0.00875*	0.00853*
	(0.00004)	(0.00008)	(0.00008)	(0.00007)
Long-term infrequent transactor	0.04508	0.0064*	0.00469*	0.00761*
	(0.00006)	(0.00011)	(0.0001)	(0.0001)
New miner	0.00389	0.00056*	0.00073*	0.00032*
	(0.00002)	(0.00003)	(0.00003)	(0.00003)
Old miner	0.00485	0.00013*	-0.00029*	0.00007
	(0.00002)	(0.00003)	(0.00003)	(0.00004)
One-time user	0.36433	0.03756*	0.04276*	0.05979*
	(0.00016)	(0.00032)	(0.00031)	(0.00032)
Other Known Entity	0.00528	-0.00058*	-0.00057*	-0.0008*
	(0.00002)	(0.00004)	(0.00004)	(0.00004)
Short-term user	0.07383	0.01325*	0.01298*	0.01293*
	(0.00008)	(0.00015)	(0.00014)	(0.00015)

We record the fraction of entities in the Americas that belong to each user type, then record the difference between that percentage and the percentage of entities in each of the other regions that belong to the same user type

* The asterisk denotes a result that is statistically significant at the 95% level.

8 Online Appendix

8.1 Data Appendix

8.1.1 Data Collection

The Bitcoin blockchain was downloaded from http://blockchain.info using the site's API. All data on the timing of transactions comes from http://blockchain.info's first observation of the transaction. This study used blockchain data through block 381500, which http://blockchain.info has dated November 30, 2015.

Information on the geographic location of Bitcoin actors associated with specific Bitcoin addresses was collected by scraping user profiles on http://Bitcointalk.org. We collected data on users who listed a location in their profile or who posted in the Other languages/locations section of the website. We associated users who posted in multiple language/location forums on the site with the location they posted to most frequently. The collected geographic data was cleaned manually to eliminate invalid places. The scripts to collect this data were last run at the end of November 2015 and collected 4838 unique Bitcoin addresses.

Data on the exchange rate of Bitcoin was published by New Liberty Standard on October 5 2009, at a rate of \$1 = 1309.03 BTC. Between October 5 2009 and May 21 2010 we calculated the daily price based on an exponential growth rate of 0.7364%, then on May 22 2010 the first documented purchase of a good was recorded for 10,000 BTC with a qouted value of \$41. From May 23 2010 to July 11 2010 we calculated the daily price based on an exponential growth rate of 1.3193%, then the value of Bitcoin rose 10x over a five day period between July 12 2010 and July 17 2010. From July 18 to present we collect the average daily price from http://www.coindesk.com/price/.

Data on Bitcoin addresses associated with specific companies, exchanges, pools, gambling sites, and other labeled categories was collected from http://www.walletexplorer.com, https://blockchain.info/tags, and Meiklejohn et al. [2013] (http://cseweb.ucsd.edu/ smeiklejohn/files/imc13.pdf). The script to collect data from Wallet Explorer was last run on April 15 2015 and collected a total 3,908,092 addresses. The data from blockchain.info was last collected on February 23 2016 and consist of all Verified addresses on the Tags page of the website, and there are 1138 addresses from the Meiklejohn et al. [2013] paper. Additional information on a small number of Bitcoin exchanges was collected by executing transactions with the sites.

8.1.2 Data Cleaning and Formatting

We extracted transaction-level data from each block. Transactions missing an address, such as those resulting from mining (no input address) or storing messages (no output address), were removed from their respective dataset. We then augment the transaction data by labeling specific transactions and addresses of interest.

Bitcoin users may engage in Mixing transactions, where multiple individual transactions are bundled together in a single transaction, in order to obscure the flow of Bitcoin between addresses. Transactions with both four or more input addresses and four or more output addresses were labeled

as Mixing transactions.

The Bitcoin associated with a single address must be used in full as an input to a transaction. When users want to transfer less than the full amount to another Bitcoin address, change from the transaction appears as a new output address in the transaction. We identify these change addresses by: 1) Whether one of the output values (amount of money sent to the output in BTC) in a two-output transaction has at least three more significant decimal places than the other output value, indicating that it is most likely the “change" from the transaction and the associated address is the change address, or 2) Whether one of the output addresses in a transaction has only received money once, indicating that it is probably a change address

8.1.3 Bitcoin Entities

Bitcoin users usually have multiple Bitcoin addresses. Ignoring mixing transactions, we group Bitcoin addresses observed in the transaction data into entities by associating addresses that were used together as inputs to a transaction together using transitive relations across transactions. So if A and B are input addresses in one transaction, and A and C are input addresses in a different transaction, addresses A, B, and C are all associated with the same entity. The entities are then grouped and consolidated in both inputs and outputs tables by replacing the entity associated with the change address with the entity associated with the input.

We then categorize entities into 10 different types: Miners new, old, Exchanges, Other Known Entity, One-Time User, Short-Term Transactor, Long-Term Frequent Transactor, Long-Term Infrequent Transactor, Investor, and Mixer.

8.1.4 Geographic Assignments

We were able to associate a country with 2995 Bitcoin entities using data collected from http://Bitcointalk.org. Since the country assignments originated at the address level, it was possible for entities to be associated with multiple countries. In cases where an entity was assigned multiple countries, the entity was labeled the country which occured most frequently. There were 137 entities that were labeled multiple countries an equal number of times, which were removed from the dataset when training a model to classify the entire network, resulting in 2858 entities.

We grouped the countries into eight regions: Asia Pacifc, Australia and New Zealand, Europe, Former USSR, Latin America, Middle East and Africa, North America, and South Asia. We then grouped these eight regions into four mega regions based on their time zones: Americas, Europe, Asia, and Eastern Latin America.

These 2858 entities were then used to train a random forest model to probabilistically classify the remainder of the network into the four mega regions. The most predictive features were those relating to the time of day an entity most frequently engaged in Bitcoin transactions, along with variables indicating whether an entity had transacted with specific exchanges. Also helpful were features we created using the addresses collected from walletexplorer.com (wallets and exchanges only). We labeled each of these address' associated entity a country and region based on their

interactions with the 2858 geographically labeled entities collected from Bitcointalk.org.

8.2 Regional Classification

8.2.1 General Information

Table 13: Mapping from Countries to Regions used in Geographic Classification

Country	Region
Algeria	Europe
Amsterdam	Europe
Argentina	Eastern Latin America
Australia	Asia
Austria	Europe
Bahrain	Europe
Bangladesh	Asia
Belarus	Asia
Belgium	Europe
Bermuda	Americas
Bosnia and Herzegovina	Europe
Botswana	Europe
Brazil	Eastern Latin America
Bulgaria	Europe
Cambodia	Asia
Canada	Americas
Caribbean	Americas
Chile	Eastern Latin America
China	Asia
Colombia	Americas
Costa Rica	Americas
Croatia	Europe
Cyprus	Europe
Czech Republic	Europe
Denmark	Europe
Dominican Republic	Americas
Ecuador	Americas
Egypt	Europe
Estonia	Asia
Ethiopia	Europe
Finland	Europe
France	Europe

French Polynesia	Asia
Germany	Europe
Greece	Europe
Guatemala	Americas
Hong Kong	Asia
Hungary	Europe
Iceland	Europe
India	Asia
Indonesia	Asia
Iran	Europe
Iraq	Europe
Ireland	Europe
Israel	Europe
Italy	Europe
Ivory Coast	Europe
Japan	Asia
Kazakhstan	Asia
Kenya	Europe
Kuwait	Europe
Laos	Asia
Latvia	Asia
Lebanon	Europe
Liberia	Europe
Lithuania	Europe
Macedonia	Europe
Malaysia	Asia
Malta	Europe
Mexico	Americas
Moldova	Asia
Mongolia	Asia
Montenegro	Europe
Morocco	Europe
Myanmar	Asia
Namibia	Europe
Nepal	Asia
Netherlands	Europe
New Zealand	Asia
Nicaragua	Americas
Norway	Europe

Pakistan	Asia
Palestine	Europe
Panama	Americas
Paraguay	Eastern Latin America
Peru	Americas
Philippines	Asia
Poland	Europe
Portugal	Europe
Romania	Europe
Russia	Asia
Saudi Arabia	Europe
Serbia	Europe
Singapore	Asia
Slovakia	Europe
Slovenia	Europe
South Africa	Europe
South Korea	Asia
Spain	Europe
Sri Lanka	Asia
Suriname	Americas
Sweden	Europe
Switzerland	Europe
Taiwan	Asia
Thailand	Asia
Tunisia	Europe
Turkey	Europe
UAE	Europe
Ukraine	Asia
United Kingdom	Europe
United States of America	Americas
Uruguay	Eastern Latin America
Uzbekistan	Asia
Venezuela	Eastern Latin America
Vietnam	Asia
Yemen	Europe
Zimbabwe	Europe

8.2.2 Random Forest Summary and Analysis Tables: Best model

	Americas	Europe	Asia	E.LatAm
Country where Most Number of Transactions	USA	USA	USA	USA
Country where Most Value Transacted	USA	USA	USA	USA

Table 14: Mode of Best Factor Variables used in Random Forest Classification

	Americas	Europe	Asia	E.LatAm
Most Active Hour of a Weekday for Outputs	15	15	15	15
Most Acitve Hour of a Weekend for Outputs	23	23	10	23
Second Most Active Weekday Hour for Outputs	16	16	16	17
Second Most Active Weekend Hour for Outputs	2	9	9	16
Third Most Active Weekday Hour for Outputs	22	16	11	18
Third Most Active Weekend Hour for Outputs	2	18	5	2
Most Active Part of a Weekday for Outputs	1	5	5	5
Second Most Active Part of a Weekday for Outputs	6	6	3	1
Third Most Active Part of a Weekend for Outputs	7	5	4	6
Most Active Hour of a Weekday for Inputs & Outputs	2	15	15	15
Most Active Hour of a Weekend for Inputs & Outputs	23	23	5	17
Second Most Active Hour of a Weekday for Inputs & Outputs	2	16	15	16
Second Most Active Hour of a Weekend for Inputs & Outputs	23	16	9	16

Third Most Active Hour of a Weekday for Inputs & Outputs	20	16	15	16
Third Most Active Hour of a Weekend for Inputs & OUtputs	3	9	12	21
Most Active Part of a Weekday for Inputs & Outputs	0	5	5	5
Most Active Part of a Weekend for Inputs & Outputs	7	7	4	5
Second Most Active Part of a Weekday for Inputs & Outputs	6	6	4	1
Third Most Active Part of a Weekday for Inputs & Outputs	7	7	4	7
Most Active Hour of any day for Outputs	23	15	15	15
Most Active Part of any day for Output	7	5	5	5
Most Active Hour of any day for Inputs & Outputs	23	15	15	15
Most Active Part of any day for Inputs & Outputs	0	5	5	5
Most Active Hour of a Weekday for Inputs	23	20	14	16
Most Active Hour of a Weekend for Inputs	1	12	5	17
Second Most Active Hour of a Weekday for Inputs	3	17	14	20
Second Most Active Hour of a Weekend for Inputs	21	15	14	20
Third Most Active Hour of a Weekday for Inputs	20	17	7	21
Most Active Part of a Weekday for Inputs	6	6	4	5
Second Most Active Part of a Weekday for Inputs	7	4	4	0

Most Active Hour of any day for Inputs	1	15	14	21
Most Active Part of any day for Inputs	0	6	4	6

Table 15: Mode of Best Time-Related Variables used in Random Forest Classification

	Americas	Europe	Asia	E.LatAm
First Seen, Number of Days Since Jan 1, 1998	1674.73	1680.30	1765.80	1690.09
Last Seen, Number of Days Since Jan 1, 1998	2064.29	2087.07	2115.88	2069.39
Percent of Transactions with Exchanges	0.20	0.24	0.23	0.22
Number of Counterparts from Known Entities	240.72	172.02	253.21	183.57
Number of days Active	441.13	454.24	401.03	416.30
Max Balance	19114.23	8840.87	13044.67	10479.58
Final USD Balance	2440.19	516.91	678.65	505.22
Number of Entites	73583.58	36425.39	115124.76	173228.54
Transaction Count of Inputs & Outputs in Most Active Hour of a Weekday	1.83	2.10	2.27	2.13
Most Active Hour of any day for Outputs	12.29	12.48	12.02	12.57
Most Active Part of any day for Outputs	3.77	3.84	3.68	3.87
Most Active Hour of any day for Inputs & Outputs	12.28	12.49	12.04	12.60
Most Active Part of any day for Inputs & Outputs	3.76	3.84	3.69	3.87
Third Most Active Weekday Hour for Inputs	-0.04	0.10	0.03	-0.01
Third Most Active Weekend Hour for Inputs	-0.63	-0.58	-0.60	-0.63
Indicator of transacting with Bitcoin.De	0.00	0.00	0.00	0.00
Indicator of transacting with Bitstamp	0.00	0.00	0.00	0.00

Indicator of transacting with BitTrex	0.00	0.00	0.00	0.00
Indicator of transacting with BTCChina	0.00	0.00	0.00	0.00
Indicator of transacting with BTC-E	0.00	0.00	0.00	0.00
Indicator of transacting with Cexio	0.00	0.00	0.00	0.00
Indicator of transacting with Cryptsy	0.00	0.00	0.00	0.00
Indicator of transacting with Kraken	0.00	0.00	0.00	0.00
Indicator of transacting with LocalBitcoins	0.00	0.00	0.00	0.00
Indicator of transacting with MintPal	0.00	0.00	0.00	0.00
Indicator of transacting with MtGox	0.00	0.00	0.00	0.00
Number of Entities Transacted with in Asian Pacific	0.01	0.02	0.05	0.02
Number of Entities Transacted with in Europe	0.08	0.11	0.11	0.11
Number of Transactions with Asia Pacific Entities	0.03	0.04	0.14	0.05
Number of Transactions with European Entities	0.21	0.29	0.36	0.37
Indicator of transacting with European entities	0.06	0.09	0.08	0.09
Number of Entities Transacted with in Australia	0.00	0.00	0.00	0.00
Number of Entities Transacted with in Canada	0.00	0.00	0.00	0.00
Number of Entities Transacted with in Czech Republic	0.00	0.01	0.00	0.00
Number of Entities Transacted with in France	0.01	0.01	0.01	0.01
Number of Entities Transacted with in Indonesia	0.02	0.03	0.06	0.07

Number of Entities Transacted with in Malaysia	0.00	0.00	0.00	0.00
Number of Entities Transacted with in United Kingdom	0.01	0.02	0.02	0.01
Transaction Volume with Entities in Canada	0.02	0.02	0.00	0.00
Transaction Volume with Entities in China	1.31	1.39	4.55	1.16
Transaction Volume with Entities in Czech Republic	0.00	0.05	0.00	0.00
Transaction Volume with Entities in Malaysia	0.01	0.01	0.00	0.00
Transaction Volume with Entities in South Africa	0.00	0.01	0.00	0.00
Transaction Volume with Entities in Thailand	0.06	0.07	0.13	0.06
Number of Transactions with Canadian Entities	0.01	0.00	0.00	0.00
Number of Transactions with Czech Republic Entities	0.00	0.01	0.00	0.00
Number of Transactions with French Entities	0.01	0.02	0.01	0.01
Number of Transactions with Malaysian Entities	0.09	0.11	0.09	0.06
Number of Transactions with South African Entities	0.00	0.01	0.00	0.00
Number of Transactions with Thailand Entities	0.01	0.02	0.03	0.02
Indicator of transacting with Canadian Entities	0.00	0.00	0.00	0.00
Indicator of transacting with French Entities	0.01	0.01	0.01	0.01
Indicator of transacting with Indonesian Entities	0.01	0.02	0.04	0.05
Indicator of transacting with Malaysian Entities	0.00	0.00	0.00	0.00
Indicator of transacting with UK Entities	0.01	0.02	0.01	0.01

Indicator of transacting with Ransomware Entities	0.00	0.00	0.00	0.00
Transaction Volume with Europe	87.09	37.14	14.34	23.75
Number of Distinct Brazilian Entities Transacted with	0.01	0.02	0.03	0.39
Transaction Volume with Branzilian Entities	0.55	0.21	0.20	26.17
Transaction Count with Brazilian Entities	0.13	0.26	0.26	16.95
Indicator of transacting with a Brazilian Entity	0.01	0.01	0.02	0.30
Transaction Count with Europe	86.85	36.94	14.12	23.56

Table 16: Average Value for Each Cluster for Best Variables used in Random Forest Classification

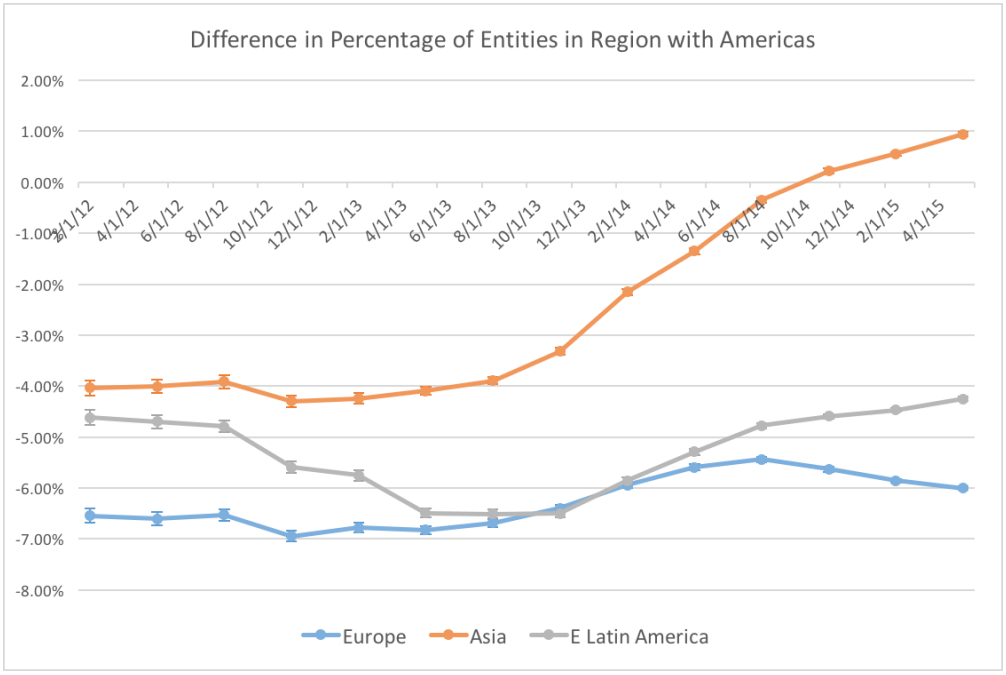


Figure 17: Difference in percentage of entities in each region (from clean model) with nonzero balance over time.

Table 17: Network Distance from Entity to Industry

Industry	America	Europe Difference	Asia Difference	E Latin America Difference
Unknown	2.97	-0.159	-0.053	0.109
	(0.234)	(0.381)	(0.369)	(0.413)
Gambling	2.771	-0.153	-0.057	0.103
	(0.235)	(0.384)	(0.372)	(0.415)
Exchange	2.57	-0.169	-0.05	0.103
	(0.234)	(0.383)	(0.369)	(0.414)
Contraband	2.797	-0.161	-0.036	0.119
	(0.235)	(0.386)	(0.37)	(0.413)
Mining	3.012	-0.154	-0.041	0.116
	(0.234)	(0.384)	(0.371)	(0.413)
Payments	3.017	-0.153	-0.037	0.12
	(0.234)	(0.383)	(0.369)	(0.414)
Wallet	3.062	-0.159	-0.046	0.113
	(0.234)	(0.382)	(0.37)	(0.414)
Laundry	3.13	-0.162	-0.049	0.11
	(0.235)	(0.384)	(0.37)	(0.414)
Giveaway	3.305	-0.159	-0.056	0.109
	(0.235)	(0.383)	(0.372)	(0.415)
Ad Network	3.148	-0.158	-0.049	0.112
	(0.234)	(0.382)	(0.37)	(0.413)
Merchant	3.258	-0.158	-0.04	0.12
	(0.233)	(0.381)	(0.368)	(0.412)
Investing	3.178	-0.161	-0.054	0.111
	(0.234)	(0.381)	(0.369)	(0.413)
Ransomware	3.438	-0.16	-0.017	0.133
	(0.235)	(0.384)	(0.369)	(0.414)

We measure network distance between the Americas and each Industry, then record the raw difference between the average distance to a given industry from a given region and the distance to the industry from the Americas

None of the differences observed between regions in this table are statistically significant at the 95% level

8.2.3 Clean Model Analyses

As described in the text, the "Clean Model" is an alternative to the regional classification model presented in the text, but where we exclude all features that are constructed using clusters (which were derived from the community detection algorithm). The Clean Model has much greater classification error. We present the results from this model for robustness.

Table 18: Distribution of Regions Predicted by Clean Random Forest Model

	Americas	Europe	Asia	E Latin America
Training Dataset	41.183%	34.990%	20.294%	3.534%
Predicted for Full Dataset	25.034%	22.547%	27.850%	24.569%
Predicted - Training	-16.149%	-12.442%	7.556%	21.035%

We compare the percentage of entities in each region in the training dataset with the percentage of entities assigned to each region by the clean random forest algorithm across the entire dataset, recording the difference between the two for each region.

Table 19: Cross-Regional Flow
Clean Model

	Region 2			
Region 1	Americas	Europe	Asia	E Latin America
Americas	-1.70699%*	-1.52014%*	-2.18162%*	-0.71283%
	(0.44624%)	(0.57293%)	(0.77542%)	(0.6837%)
Europe		0.44785%	0.52989%	1.64376%
		(0.46793%)	(0.77258%)	(0.87102%)
Asia			0.26087%	1.57889%*
			(0.88478%)	(0.76115%)
E Latin America				1.66034%*
				(0.81208%)

We first calculate the (counterfactual) share of transaction volume that would occur across regions if the region labels were assigned randomly to entities in proportion to the assignment proportions predicted by the clean random forest algorithm. We then calculate the actual share of transaction volume according to the assignments from 100 draws from the posterior distribution from the clean random forest algorithm. We report the difference between the mean of those draws and the counterfactual mean from random label assignment. We also report the standard deviation from the 100 draws.

* The asterisk denotes a result that is statistically significant at the 95% level.

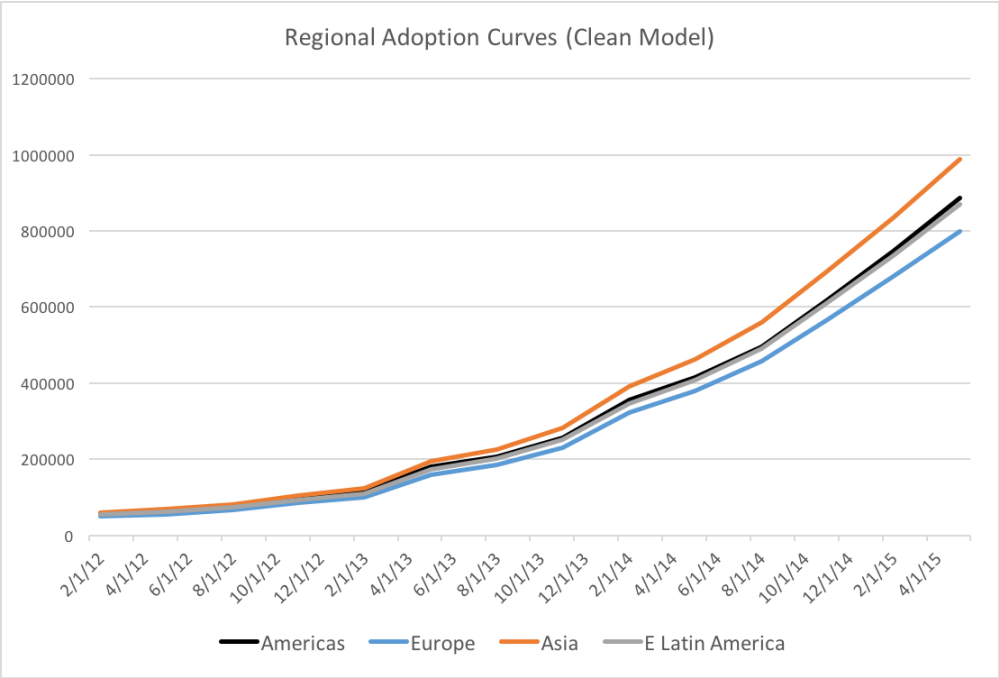


Figure 18: Number of entities in each region (from clean model) with nonzero balance over time.

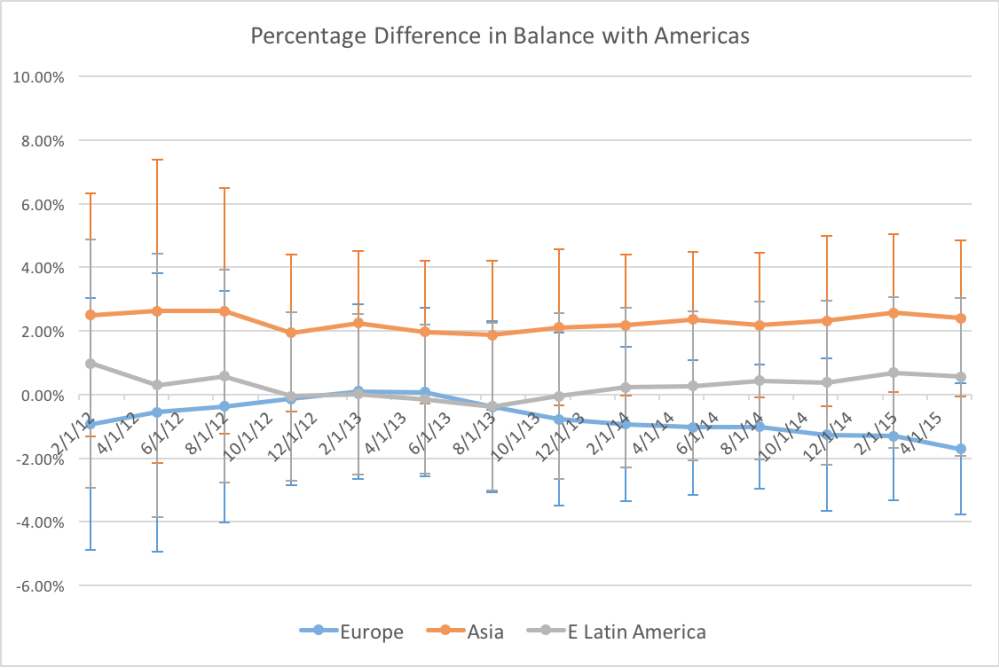


Figure 19: Percentage difference between total balance of all the entities in each region (assigned by the clean model) and the Americas ((Balance in Region - Balance in Americas)/Balance in Americas)*100%.

Table 20: Industry-Region Flow
Clean Model

Industry	Americas	Europe Difference	Asia Difference	E Latin America Difference
Unknown	0.34691	0.00143	0.03446	0.04291
	(0.06662)	(0.10428)	(0.11025)	(0.10351)
Gambling	0.07696	-0.01681	-0.01962	-0.01481
	(0.01323)	(0.01924)	(0.01957)	(0.01747)
Exchange	0.46802	0.01991	-0.00146	-0.01378
	(0.05311)	(0.07984)	(0.08069)	(0.07869)
Contraband	0.04119	-0.00275	-0.007	-0.00615
	(0.00306)	(0.00453)	(0.00387)	(0.0052)
Mining	0.02912	0.00096	-0.00091	-0.00517
	(0.00447)	(0.00658)	(0.00692)	(0.00615)
Payments	0.01489	-0.0001	-0.00118	-0.00221
	(0.00462)	(0.00661)	(0.00711)	(0.00725)
Wallet	0.01421	-0.00227	-0.00338	-0.00122
	(0.00635)	(0.00952)	(0.00847)	(0.00932)
Laundry	0.00697	-0.00031	-0.00061	0.00052
	(0.00199)	(0.00355)	(0.00316)	(0.00314)
Giveaway	0	0	0	0
	(0)	(0)	(0)	(0)
Ad Network	0.00008	0	0	0.00001
	(0.00001)	(0.00002)	(0.00002)	(0.00002)
Merchant	0.00013	-0.00001	-0.00003	-0.00002
	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Investing	0.0015	-0.00006	-0.00026	-0.0001
	(0.00021)	(0.0003)	(0.00031)	(0.00029)
Ransomware	0.00002	0	0	0.00001
	(0.00002)	(0.00002)	(0.00003)	(0.00003)

We record the percentage of bitcoin flow involving the Americas that is with each industry. Then, for each other region we record the difference between the percentage of flow with a given industry for the given region and for the Americas
None of the differences observed between regions in this table are statistically significant at the 95% level

Table 21: Network Distance from Entity to Industry
Weighted by Entity Transaction Volume
Clean Model

Industry	Americas	Europe Difference	Asia Difference	E Latin America Difference
Unknown	1.91	-0.049	-0.043	-0.05
	(0.019)	(0.027)	(0.027)	(0.032)
Gambling	1.973	-0.043	-0.021	-0.058
	(0.018)	(0.026)	(0.026)	(0.031)
Exchange	1.459	-0.052*	-0.025	-0.03
	(0.016)	(0.025)	(0.024)	(0.028)
Contraband	1.938	-0.055	-0.019	-0.056
	(0.019)	(0.029)	(0.028)	(0.031)
Mining	2.086	-0.05*	-0.031	-0.065*
	(0.017)	(0.025)	(0.026)	(0.03)
Payments	2.108	-0.054*	-0.023	-0.064*
	(0.017)	(0.025)	(0.026)	(0.028)
Wallet	2.13	-0.043	-0.023	-0.058*
	(0.018)	(0.027)	(0.026)	(0.029)
Laundry	2.204	-0.05	-0.024	-0.057*
	(0.017)	(0.027)	(0.024)	(0.028)
Giveaway	2.375	-0.049	-0.026	-0.044
	(0.018)	(0.027)	(0.025)	(0.029)
Ad Network	2.249	-0.042	-0.027	-0.053*
	(0.016)	(0.025)	(0.025)	(0.026)
Merchant	2.308	-0.044	-0.02	-0.045
	(0.016)	(0.026)	(0.023)	(0.027)
Investing	2.239	-0.049	-0.029	-0.066*
	(0.016)	(0.026)	(0.024)	(0.027)
Ransomware	2.437	-0.048	-0.019	-0.037
	(0.017)	(0.026)	(0.026)	(0.028)

We measure network distance between the Americas and each Industry, then record the raw difference between the average distance to a given industry from a given region and the distance to the industry from the Americas, weighting averages by their total transaction volume with other entities

* The asterisk denotes a result that is statistically significant at the 95% level.

Table 22: Fraction of Entities of each User Type by Region
Clean Model

UserType	Americas	Europe Difference	Asia Difference	E Latin America Difference
Mixing	0.40991	-0.04278*	-0.01518*	-0.01718*
	(0.00022)	(0.0004)	(0.00031)	(0.00033)
Exchange	0.00217	0.00002	0.00007*	0.00004
	(0.00001)	(0.00002)	(0.00003)	(0.00003)
Investor	0.03499	0.005*	0.00355*	0.0031*
	(0.00006)	(0.00009)	(0.00011)	(0.0001)
Long-term frequent transactor	0.0212	0.00747*	0.00483*	0.00343*
	(0.00004)	(0.00008)	(0.00007)	(0.00008)
Long-term infrequent transactor	0.04778	0.00361*	0.00151*	0.00146*
	(0.00007)	(0.00013)	(0.00012)	(0.00012)
New miner	0.00404	0.00041*	0.00037*	0.0002*
	(0.00002)	(0.00003)	(0.00004)	(0.00003)
Old miner	0.00472	0.00026*	0.00015*	0
	(0.00002)	(0.00004)	(0.00003)	(0.00004)
One-time user	0.3921	0.01482*	0.00047	0.00469*
	(0.0002)	(0.00032)	(0.00035)	(0.00033)
Other Known Entity	0.0049	-0.00019*	-0.0001*	-0.00001
	(0.00002)	(0.00004)	(0.00004)	(0.00004)
Short-term user	0.07818	0.01137*	0.00433*	0.00427*
	(0.00007)	(0.00014)	(0.00012)	(0.00015)

We record the fraction of entities in the Americas that belong to each user type, then record the difference between that percentage and the percentage of entities in each of the other regions that belong to the same user type

* The asterisk denotes a result that is statistically significant at the 95% level.

Table 23: Network Distance from Entity to Industry
Clean Model

Industry	America	Europe Difference	Asia Difference	E Latin America Difference
Unknown	3.04	-0.155	-0.078	-0.156
	(0.253)	(0.435)	(0.375)	(0.392)
Gambling	2.841	-0.157	-0.076	-0.158
	(0.253)	(0.436)	(0.374)	(0.393)
Exchange	2.642	-0.168	-0.079	-0.162
	(0.253)	(0.434)	(0.376)	(0.393)
Contraband	2.872	-0.158	-0.069	-0.156
	(0.253)	(0.436)	(0.374)	(0.392)
Mining	3.087	-0.155	-0.075	-0.155
	(0.253)	(0.435)	(0.375)	(0.394)
Payments	3.095	-0.158	-0.072	-0.158
	(0.254)	(0.437)	(0.375)	(0.393)
Wallet	3.134	-0.154	-0.077	-0.154
	(0.252)	(0.434)	(0.374)	(0.392)
Laundry	3.201	-0.158	-0.075	-0.157
	(0.253)	(0.436)	(0.375)	(0.393)
Giveaway	3.38	-0.166	-0.083	-0.164
	(0.254)	(0.435)	(0.377)	(0.394)
Ad Network	3.224	-0.16	-0.085	-0.16
	(0.253)	(0.434)	(0.376)	(0.393)
Merchant	3.335	-0.157	-0.076	-0.158
	(0.254)	(0.436)	(0.375)	(0.394)
Investing	3.248	-0.159	-0.078	-0.157
	(0.253)	(0.435)	(0.375)	(0.393)
Ransomware	3.518	-0.155	-0.066	-0.152
	(0.253)	(0.434)	(0.374)	(0.393)

We measure network distance between the Americas and each Industry, then record the raw difference between the average distance to a given industry from a given region and the distance to the industry from the Americas
None of the differences observed between regions in this table are statistically significant at the 95% level

	Predicted				
Actual	Americas	Europe	Asia	E. Latin America	Class Error
Americas	38.912%	19.541%	25.913%	15.633%	0.611
Europe	13.900%	43.100%	31.800%	11.200%	0.569
Asia	10.517%	19.655%	60.172%	9.655%	0.398
E Latin America	19.802%	13.861%	22.772%	43.564%	0.564

Table 24: Random Forest OOB Accuracy using Clean Set Variables

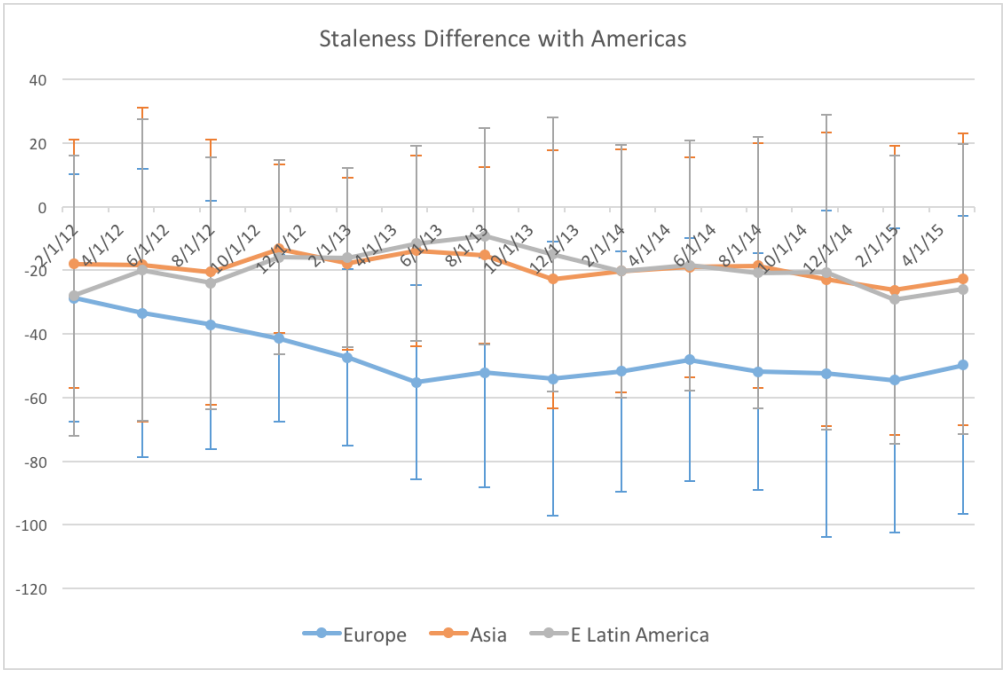


Figure 20: Difference between average staleness of Bitcoins held by entities in each region (assigned by the clean model) and the Americas (Staleness in Country - Staleness in Americas).

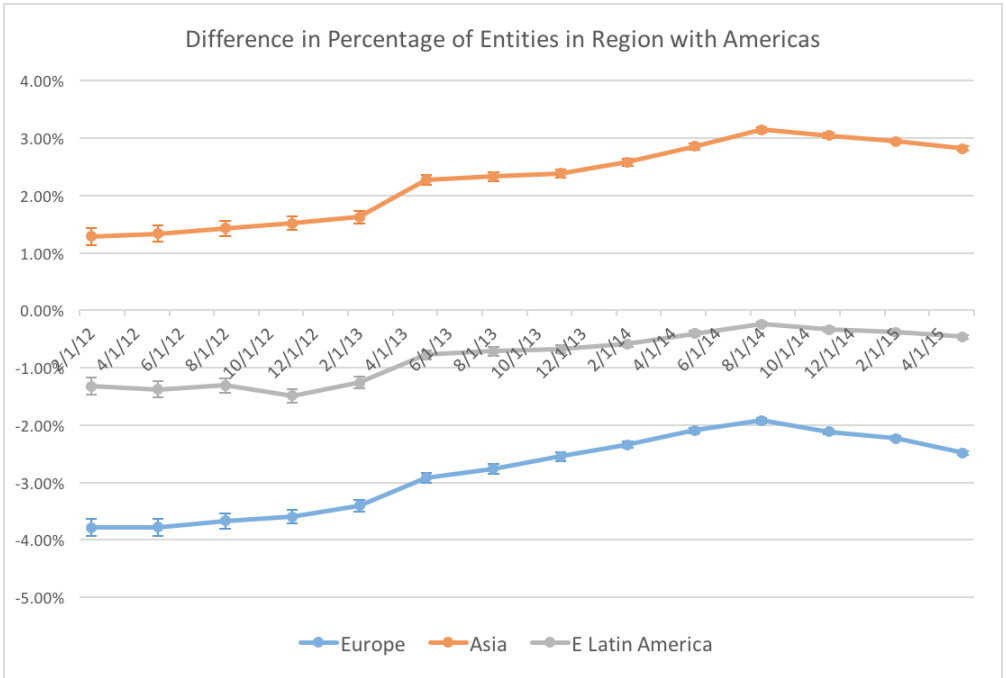


Figure 21: Difference in percentage of entities in each region (from clean model) with nonzero balance over time.