# BITCOIN AND ETHEREUM: EMPIRICAL EVIDENCE ON NODE DISTRIBUTION

A THESIS

Presented to

The Faculty of the College of Arts and Sciences

The University of Colorado at Boulder

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Arts

By

Keaton Brown

April 2018

Advised by: Professor Martin Boileau

BITCOIN AND ETHEREUM: EMPIRICAL EVIDENCE IN NODE DISTRIBUTION

Keaton Brown

April 2018

Economics

## Abstract

With the explosive growth in cryptocurrencies over the last couple of years, the cost of *mining* these technologies (the process through which users devote CPU power to operate the underlying blockchains) have similarly exploded. This paper examines one overarching question regarding this issue – what factor or factors explain the geographic distribution of cryptocurrency nodes (mining operations) across the world? In exploring this question, this research considers electricity price, internet access, Tor network relays, and others. Using node distribution data for Bitcoin and Ethereum – the two largest cryptocurrencies – this paper analyzes cross-sectional and panel data regression models, and establishes that electricity price has not played a significant role in this distribution up to this point, and concludes that the historical association between Tor relays and Bitcoin use has had a much greater impact. Lastly, this paper discusses the broader implications of its findings, and the potential areas of research for further understanding of this field.

# ACKNOWLEDGEMENTS

In the initial development of this paper, great insight and advice as to potential literature and data sources to draw off of were suggested by Giri Subramaniam, a graduate assistant at the Federal Reserve of Boston inside the Consumer Payments Research Division.

Throughout the paper, in order to collect all data for Bitcoin and Ether nodes, Karl Roos, a friend and colleague, provided immense technical assistance in writing all necessary code to scrape the data used in this paper. Without his assistance, this research would not have been possible.

This paper thanks both Giri and Karl for their immense help in developing and conducting the research examined in this paper.

# TABLE OF CONTENTS

INTRODUCTION

Within the months leading up to the publication of this paper, numerous reports have taken into question whether – due to the astronomical energy usage and cost of mining Bitcoin – cryptocurrencies are a sustainable technology. Using electricity price data and cryptocurrency node data across countries, this paper establishes the extent to which electricity price influences the prevalence of cryptocurrencies nodes within a country. Originally, this paper intended to answer one question: *How does the effect of electricity price on node count depend on the level of internet access within a given country?* However, as we will see later, electricity prices do not significantly explain the presence of nodes. Thus, a search for significant drivers in node count led to a second research question: *How do the number of Tor network relays affect the number of cryptocurrency nodes within a given country*? Both research questions provide interesting insights into the geographic distribution of Bitcoin and Ethereum nodes.

The issue of electricity cost in Bitcoin mining has been a recurring and persistent point of debate in cryptoeconomic discourse. To place the scale of Bitcoin electricity use in perspective, a recent report found that the one-year electricity consumption of the Bitcoin network was, "…commensurate with CO2 emissions of 20 megatonnes – or roughly 1m transatlantic flights." (Hern, 2018) With annual electricity usage encompassing the total usage of some G20 countries, the feasibility of the Bitcoin network – both in terms of cost and environmental degradation – has come into question.

Despite the looming and inevitable electricity problem that exists within the cryptoeconomic sphere, the literature regarding this issue is sparse, as most literature has delved into valuation methods for cryptocurrencies. Several notable papers, however, provide a baseline for this research. First, Athey et. al. (2016) constructs a theoretical adoption model of Bitcoin, and touches somewhat on geographic distribution and use. Second, Hayes (2017) examines electricity price with regards to cryptocurrencies. The paper, however, considers electricity price solely as a valuation method for price predictions. As research in this field is relatively scarce, this paper will contribute to the cryptoeconomic field in three ways: first, it establishes that electricity price, up to this point, has played an insignificant role in the geographic distribution of Bitcoin and Ether nodes. Second, it posits that the most significant driver in node distribution has been Tor network relays, which suggests that distribution up to this point has had much more to do with historical ties than rational economic decisions. Lastly, drawing upon both initial findings, this paper offers insight as to when or whether electricity price might play a role in distribution in the future.

## BACKGROUND DEFINITIONS

As cryptoeconomics is a very new field, a background section separate from an introduction was deemed necessary to define key words and ideas mentioned throughout this paper.

Cryptocurrency: “a digital currency in which encryption techniques are used to regulate the generation of units of currency and verify the transfer.” (Oxford Dictionary, 2018)

Bitcoin: the largest by market capitalization and generally deemed the first cryptocurrency, it was created in 2008 by a person or persons using the pseudonym “Satoshi Nakamoto”. (Nakamoto, 2008)

Ether: the second largest cryptocurrency by market capitalization, Ether was created in 2015 with several unique features separate from Bitcoin. Ether is the currency of Ethereum, the underlying blockchain.

Blockchain: the underlying technology on which cryptocurrencies operate – "a digitized, decentralized, public ledger of all cryptocurrency transactions." (Investopedia, 2018) Bitcoin and Ether operate on separate blockchains.

Mining: the process through which transactions are verified on the blockchain through the operation of "nodes", where individuals devote CPU power to verify transactions, and in turn receive transactional fees and "mine" additional units of the associated cryptocurrency as they are created.

Tor: an anonymized internet browser developed by the US military in the 1990s, where the identity and location of individuals using the browser are masked.

Tor relay: essentially a "node" of the Tor network. Thousands of Tor relays across the world anonymize the IP address of an individual using the Tor browser by bouncing the IP address across a randomized selection of relays using high levels of cryptography.

Additional terms will be presented as they arise in the research, and the terms will be covered more extensively where necessary, but a general understanding of the above terms helps present an idea of the general cryptoeconomic ecosystem.

## LITERATURE REVIEW

Within the field of cryptoeconomic research, literature can be divided into four broad categories: financial analysis, political and regulatory affairs, technology flaws and applications, and macroeconomic analysis. The first category, financial analysis, deals mostly with research involving pricing predictions, although other financial characteristics of cryptocurrency – especially volatility – have been extensively examined. Political and regulatory affairs are primarily related with the intersection of government and cryptocurrencies/blockchain technology – how taxing structures might look and whether governments might oppose digital currencies. Technology research in this field delves into some of the supposed flaws behind blockchain technology – vulnerability to hacks, for instance – and the different applications of blockchain technology. Lastly, macroeconomic research into cryptocurrencies deals with a much larger scope – macroeconomic implications of cryptoeconomic growth including impacts on international monetary flows, effects on workforces and productivity, and the different characteristics of large-scale use for cryptocurrencies.

Predicting value in cryptocurrencies is the most extensively-studied area in this field. Studies have concluded that traditional pricing mechanisms for other assets – equities, for instance – are vastly different than predictive algorithms for cryptocurrencies, as digital currencies don't publish financial statements upon which to base value. Because of this, research in this field has delved into several different predictors for cryptocurrency price. One such paper structured a pricing model based on mining, in which the model could predict (with some degree of success) the price of 66 different cryptocurrencies using "the level of competition in the network of producers, the rate of unit production, and the difficulty of algorithm used to "mine" for the cryptocurrency." (Hayes, 2017) Further valuation models have taken other factors into account, such as the volume of cryptoeconomic-related media mentions (Polasik et al., 2015). Along with pricing and valuation models, other papers have studied in-depth the volatility of cryptocurrency prices. For instance, the volume of cryptocurrency transacted as a predictor of volatility (Balcilar et. al, 2016) and a GARCH volatility analysis as a means to classify Bitcoin in an asset class (Dyhrberg, 2015) are two preeminent papers in the field.

The potential flaws of blockchain technology were first examined in the paper which created this entire field, "Bitcoin: a peer-to-peer electronic cash system" (Nakamoto, 2008). (The author (or authors) of this paper, Satoshi Nakamoto, used a pseudonym to develop what is now Bitcoin). Other potential flaws and specific applications of blockchain technology (such as smart contracts and crypoeconomic financial transactions) and proposed future applications (such as digital identities) have been examined. (Pilkington, 2015)

Regulatory research into cryptocurrencies is a quickly-expanding field. Digital asset exchange regulations (Pieters & Vivanco, 2017), along with effects of these technologies on government functions such as monetary policy and central bank seigniorage (Committee on Payments and Market Infrastructures, 2015), have been two areas of thorough examination.  Tax implications have been a controversial area within cryptoeconomic research, but the current state of government cryptoeconomic tax policies, along with routes in which these policies may follow, have been examined by Akins et. al, 2015. Since this paper, however, a variety of new regulatory legislation has been introduced, so further research into this field is necessary.

Macroeconomic analysis primarily explores the growth of cryptocurrencies and blockchain technology and the methods in which these technologies are being used. This is also the field to which this paper contributes. Adoption of cryptocurrencies – both empirically and theoretically – has been extensively covered (Athey et. al, 2016), and closely relates to the research conducted in this paper. Athey's paper further examines aspects such as geographic cryptocurrency use, which is further explored through the research in this paper. User characteristics have also been analyzed through the analysis of Google Trends data (Yelowitz & Wilson, 2015). Along with uses, this field has developed research into the economic growth these technologies have – and could – facilitate. For instance, the use of cryptocurrencies in developing countries for stability and regional payments has been explored (Magee, 2015). Furthermore, the growth of cryptocurrencies in different functions – for example, payments and wallets (virtual “bank accounts”) – along with the growing number of users and even careers within the field, has been explored in great detail. (Hileman & Rauchs, 2017) This paper combines aspects from each of the aforementioned papers to delve further into the examinations of adoption, use, and distribution.

Cryptoeconomics is a quickly-growing field, and finance, technology, politics, and macroeconomic analysis have been areas of extensive research, even if Bitcoin has existed only for nine years. The research conducted in this paper will likely touch on each of the four categories in varying degrees, however, macroeconomic analysis is surely the category in which this paper will contribute most extensively. Branching off some of the geographic analyses performed in papers such as Athey et. al (2016) through the utilization of blockchain node data is, by my knowledge, a first-of-its-kind look into these technologies.

## DATA

### Cross-Sectional:

Data for this research has been obtained from a variety of sources. The Bitcoin node and Ether node data are of most critical importance. To obtain Bitcoin data, Karl Roos (mentioned under the “Acknowledgements” section) provided immense technical assistance, and wrote scripts to gather this data from earn.com, a popular bitcoin website which, as a side project, created bitnodes.earn.com, from which the relevant data was scraped. Roos summarized the method for scraping this data: “Earn.com has a project aimed at cataloguing all the Bitcoin nodes in the world over time – their method of doing this is by pinging nodes on the Bitcoin network at a specific time (snapshot). They then log the IP and attach meta data (geo IP information for example). Our method for obtaining the data was to build a script that utilized the open API to first list all the available snapshots and then download them, they currently publish the previous 180 days with a granularity of roughly every 5-6 minutes which means a total of about 46,000 files." From these 46,000 files, 200 were sampled over the 180-day time frame and compiled to construct a list of average number of nodes per country. Similarly, for Ether data, Roos wrote another script, this time compiling data from ethernodes.org, another blockchain data aggregator. Due to the lack of variation in node count across the 180-day timeframe analyzed for Bitcoin, only the most recent snap shot was scraped for Ether data (around mid-January 2018). (Figures 1 & 2 display a geographical density map of the current distribution of nodes.)

Once the Bitcoin and Ether data were complete, data for other variables were relatively simple to collect. The primary variables of interest in the initial regressions are electricity price (a measure of electricity cost, in cents, per kilowatt hour) and internet access (the proportion of people with access to the internet within a country). Electricity price data is sourced from Wikipedia, as minimal data is available. However, per the article, the price data is collected from official government sources. Internet access data, along with bank account (the proportion of a country with a bank account), real GDP (real GDP of the country in 2010 dollars), GDP growth (one-year growth of the country's GDP), and population (population of the country) data are sourced from the World Bank – these data sets are from 2016. Innovation scores (a country's Global Innovation Index) are sourced from the World Intellectual Property Association, also from 2016.

In the second set of cross-sectional regressions, the primary variable of interest is the number of Tor network relays per country. The Tor relay data is obtained from Tor Metrics, the publicly-available dataset on Tor use provided by the Tor Foundation. (Figure 3 displays a density map of the current distribution of Tor relays). The only other additional dataset to the second set of cross-sectional regressions is a liberty index which estimated – on a scale of 1 to 10 – the degree of personal liberties enjoyed in each country. This index was sourced from the Cato Institute and created in 2016. A summary of cross-sectional data, listing averages, ranges, expected signs and other relevant information is listed below:

Table 1: Cross-Sectional Data Descriptive Statistics

**Cross Sectional Data**

| **Variable** | **Unit** | **Mean** | **Min** | **Max** | **Expected Sign** | **Index Range** |
|---|---|---|---|---|---|---|
| bitNode | node | 182.05 | 2 | 3061 | N/A | N/A |
| ethNode | node | 493.82 | 6 | 8470 | N/A | N/A |
| elecPrice | cents/kwh | 16.91 | 4 | 41.8 | ( - ) | N/A |
| internet | proportion | 0.67 | 0.11 | 0.96 | ( + ) | N/A |
| torRelay | relay | 37.2K | 381 | 489.4K | ( + ) | N/A |
| innovate | (index) | 43.45 | 21.9 | 63.8 | ( + ) | 0 to 70 |
| gdpReal | 2010 USD | 1.240T | 7.33B | 16.9T | ( + ) | N/A |
| gdpGrowth | % | 2.52 | -3.59 | 7.11 | ( + ) | N/A |
| pop | person | 101.2M | 1.950M | 1.409B | ( + ) | N/A |
| liberty | (index) | 7.5 | 5.37 | 8.88 | ( + ) | 0 to 10 |

Panel Analysis:

Data on cryptocurrency node distribution across time is *extremely* scarce. To obtain this data, an exhaustive search led to a script repository on a popular code-sourcing website, GitLab. Again, Karl Roos provided immense assistance in successfully cloning and running this code, and Bitcoin node data per country was obtained for the years 2013-2016. The Bitcoin panel dataset includes any countries which had at least one Bitcoin node for one or more years within the time frame under consideration. Unfortunately, as Bitcoin is relatively older than Ethereum (Bitcoin being founded in 2008, Ethereum in 2015), node data across time for Ether is unavailable.

Several additional variables are included in the panel set, all for the years 2013-2016. Data on democratic standing (with countries being ranked from -10 to 10) is sourced from the Polity Project, data on the degree of economic freedoms enjoyed per country is sourced from the Heritage Foundation (the index ranging from 0 to 100, with higher scores corresponding to more economic freedoms), and a corruption index is sourced from Transparency International (also from 0 to 100, with higher scores corresponding to higher levels of corruption). Lastly, an index examining the extent of money laundering in the country is later included, this index being sourced from Basel Governance (scores ranging from 1 to 10, with higher scores corresponding to higher degrees of laundering activity). A summary of relevant descriptive statistics for the panel data is listed below:

Table 2: Panel Data Descriptive Statistics

**Panel Data**

| **Variable** | **Unit** | **Mean** | **Min** | **Max** | **Expected Sign** | **Index Range** |
|---|---|---|---|---|---|---|
| torRelay | relay | 24.5K | 66 | 424.5K | ( + ) | N/A |
| internet | % | 65.81 | 12 | 97.49 | ( + ) | N/A |
| innovate | (index) | 43.65 | 23.1 | 68.3 | ( + ) | 0 to 70 |
| gdpReal | 2010 USD | 919.0B | 4.3B | 16.9B | ( + ) | N/A |
| gdpGrowth | % | 2.74 | -9.77 | 25.56 | ( + ) | N/A |
| pop | person | 72.3M | 423K | 1.38B | ( + ) | N/A |
| polity | (index) | 6.48 | -10 | 10 | ( + ) | -10 to 10 |
| econFreedom | (index) | 65.93 | 40.27 | 9.01E+01 | ( + ) | 0 to 100 |
| corruption | (index) | 53.86 | 24 | 92 | ( + ) | 0 to 100 |
| launder | (index) | 5.21 | 2.51 | 8.61 | ( + ) | 0 to 10 |

METHODOLOGY

To conduct the initial regressions, the primary variable of interest in predicting node count for both Bitcoin and Ether is an interaction term between electricity price and internet access. (Figures 4 & 5 display scatterplots of the relationship between electricity price and nodes). Through the interaction of electricity price and internet access, the initial regressions seek to examine under which conditions electricity price is significant. The initial regression is as follows:

$$nodes^*_i = \beta_0 + \beta_1[elecPrice_i \ x \ internet_i] + \beta_2 elecPrice_i + \beta_3 internet_i + \beta_4 innovate_i + \beta_5 bankAccount_i + \beta_6 log(pop_i) + \beta_7 log(realGDP)_i + \beta_8 gdpGrowth_i + \varepsilon_i$$

(*where nodes refers to either Bitcoin or Ether nodes)

As mentioned previously, several variations of the initial cross-sectional regressions examining electricity price yield insignificant results. As such, two new cross-sectional regressions (one for Bitcoin and one for Ether) are constructed to find a more significant predictor of node count, this time examining Tor network relays. A Tor relay can be thought of like a Bitcoin node – individuals across the world can set up a Tor relay wherever internet is available. For the Tor network to operate successfully, many of these relays are necessary, as an individual will access the internet via Tor, and Tor will then reroute the individual's IP address (and thus location) across a randomized number and pathway of relay locations. This makes tracking the individual's activity and location incredibly difficult, almost impossible.

Tor is the primary method to access the Deep Web – a "hidden" underlayer of the internet for which pages are not indexed by search engines. As Tor provides an anonymous, encrypted way to access the internet, it is often used to access illegal marketplaces on the Deep Web, where guns, drugs, counterfeit documents, and any number of illegal goods and services are available for purchase. Many times, Bitcoin is the easiest way to pay for these goods and services, which suggests a causal relationship of Tor relays on Bitcoin nodes. The cross-sectional regression examining Tor relays is listed below:

$$nodes_i = \beta_0 + \beta_1 torRelay_i + \beta_2 elecPrice_i + \beta_3 internet_i + \beta_4 innovate_i + \beta_5 bankAccount_i + \beta_6 log(pop_i) + \beta_7 log(realGDP)_i + \beta_8 gdpGrowth_i + \beta_9 [elecPrice_i \text{ x } internet_i] + \varepsilon_i$$

After conducting the cross-sectional analysis, a panel model is constructed for Bitcoin, modeling a four-year period (2013-2016). In this analysis, the primary variable of interest is Tor relays. As will be discussed later, Tor relays significantly affects the number of nodes. Other controls are employed, including *freedom* (a ranking of socioeconomic freedoms enjoyed by countries and developed by the Cato Institute, 2016), which is included to isolate the effect of Tor relays on Bitcoin nodes. The reasoning behind this inclusion stemmed from Athey et. al.'s (2016) research, which identified three main Bitcoin user groups: libertarians, criminals, and traders. As libertarianism and Tor use can often go hand-in-hand, the freedom index is included to separate these effects.

## RESULTS

### Cross-Sectional: Electricity Price and Internet Access

In the initial phase of this paper, the significance of electricity price – dependent on a given level of internet access – is the variable of primary interest. However, neither electricity price, nor the interaction between electricity price and internet, proved significant. For example, the average proportion of internet access through all countries in this sample is 0.67. Taking the interaction term coefficient (-44.28), multiplying this by the 0.67 internet average, and adding the standalone electricity price coefficient (29.4) shows that, at this level of internet access, a one-cent increase in electricity price corresponds to a 0.26 decrease in the number of Bitcoin nodes. (See Table 3, Regression 1) The sign of the interaction term coefficient is interesting. Both electricity price and internet access have positive (non-significant) coefficients, but the effect of electricity price on node count given an increase in internet access had a negative effect, suggesting that the sign of the effect of electricity price on node count changed depending on some threshold for internet access. To calculate this threshold, we examined the derivative of the first regression with respect to electricity price, assigned *s* to represent some threshold internet value, and set the equation equal to zero (shown below).

$$0 = (-44.28)s + 29.4$$

For Bitcoin nodes, this threshold is 66.4% -- countries with internet access values above this experienced negative effects of electricity price on node count, and countries with internet access values below this experienced positive effects of electricity price on node count. More broadly speaking, this suggests that node agents in developed countries (with higher internet levels) are more inclined to be adversely responsive to increases in electricity prices, while developing countries are less likely to diminish node operations based on electricity price increases. Although electricity price and internet access had insignificant effects in this regression, the R-squared is fairly high, with 30.93% of the variation in Bitcoin nodes being explained by the regression, likely due to the fact that the innovation index is significant at the 5% level, and population is significant at the 10% level. (See Table 3, Regression 1)

Results are similar in initial cross-section Ether models. At the same average internet access level of 67%, by taking the interaction term coefficient (-204.29), multiplying this by the 0.67 internet average, and adding the standalone electricity price coefficient (121.12), we find that a one-cent increase in electricity price corresponds to a 15.75 decrease in Ether node count. (See Table 4, Regression 1) This time, the coefficient is significant, but at the lowest level (10%). Though the interaction effect is greater in the Ether model, electricity prices in the sample only range from 4 to 42 cents per kwh, so a one-cent increase represents a substantial price increase. Again, both electricity price and internet access on their own have positive effects on Ether node count, but the interaction term between the two variables had a negative effect. The threshold of internet access for the Ether regression for which positive and negative effects of electricity price are centered, calculated using the same methods as in the Bitcoin node model, is 59.3%. The adjusted R-squared in this model is 35.9%, this time with the innovation index being significant at the 1% level, and population significant at the 5% level. (See Table 4, Regression 1)

Although the interaction between electricity price and internet access met some level of statistical significance in the Ether models, this interaction is not statistically significant in the Bitcoin model, and not nearly as significant as initially expected. This statistical insignificance of electricity price led to another question – *if electricity price isn't a significant predictor of node count, what is?* The significance of the innovation index and population are, for the most part, uninteresting, as their influence is self-explanatory. More technologically-sound countries and more populous countries will likely have greater counts of both Bitcoin and Ether nodes.

Cross-Sectional: Tor Relay

There is much literature regarding Bitcoin user groups, and specifically those in the "illegal activity" group. Foley et. al. (2018) estimated that around $72 billion in illegal activity is processed with Bitcoin each year. Athey et. al (2016) similarly recognized the extensive use of Bitcoin to participate in laundering and illegal activity. The extensive literature covering Bitcoin and its relation to illegal activity prompted a second set of models, this time using Tor network relays as a predictor of Bitcoin and Ether nodes. Tor is an anonymized internet browser which masks the profile and location of an individual. Using the Tor browser, an individual can access the so-called "Dark Web" and purchase drugs, weapons, counterfeit documents, and various illegal goods and services. As discussed briefly in the background section of this paper, a Tor Relay can be thought of akin to a node – dispersed around the globe, operated by individuals, and used to operate the Tor network. The more Tor relays in the world, the more secure the Tor network. After an extensive review of the literature concerning illegal Bitcoin activity and the utilization of Tor to facilitate this activity, the second set of cross sectional regressions analyzed the effect of Tor relays on Bitcoin and Ether node count.

The results of these new models establish the significance of Tor relays in node count. To preface these results, Bitcoin node count in this regression ranged from 2 to 3,061, Ether node count from 6 to 8,740, and Tor relay count from 381 to 489,483. A 10-relay increase in Tor relays, for example, led to a 0.036 increase in the number of Bitcoin nodes in a country. The positive sign of this coefficient was expected, and the coefficient is significant at the 0.1% level. Similarly, with Ether, a ten-node increase in Tor relays corresponds to a 0.089 increase in the number of Ether nodes in a country, a value significant at the 0.1% level. Both models yield high R-squared values, with 64.09% and 64.85% of the variation in Bitcoin and Ether node count being explained by the regression. Interestingly, the positive signs on electricity price and internet access in the previous set of regression become negative in the Tor relays regressions, suggesting that the coefficient were positively biased until the inclusion of Tor relay accounted for this positive association. The innovation index is again positive and significant at the 1% level in both regressions, while population is positive and significant at the 10% level only in the Ether regression. (See Tables 3 & 4, Regression 2)

Although Tor relays proved to have a very significant role in node count, another set of regressions are modeled to validate these effects. In conducting the next set of regressions, we supposed that a significant portion of Tor users are – as literature has discussed – libertarians (see Athey et. al, 2016). Perhaps Tor relay by itself isn't a good predictor of node count, because it captures the effect of libertarians on node count. To isolate these effects, we ran another set of regressions – identical to the Tor relay models – however, this time including a country-specific liberty index developed by the CATO Institute (a libertarian think-tank). Actual data on the number or proportion of libertarians per country would have likely been more relevant, however, this data is not available. These models do not alter our original findings of the significance of Tor relay. The coefficient of Tor on Bitcoin node count changed from 0.036 to 0.037, while remaining significant at the 0.1% level. Similarly, the coefficient of Tor on Ether node count changed from 0.089 to 0.091, still remaining significant at the 0.1% level. The R-squared values increased slightly, with 67.17% and 67.86% of the variation in Bitcoin and Ether node count, respectively, being explained by the variables under consideration. (See Tables 3 & 4, Regression 3)

The initial cross-sectional results provide great insight into the distribution of Bitcoin and Ether nodes – namely, they demonstrate that electricity price doesn't matter in this distribution, or at least it hasn't mattered up to this point. Furthermore, these results demonstrate that Tor relays play a very significant role in this distribution. After the cross-sectional regressions were completed, an exhaustive search for node count data over a period of years led to the discovery of Bitcoin node data from the years 2013-2018. Unfortunately, most of the variables included in the cross-sectional regressions only contained data through 2016, and no such data was available for Ether, so a panel of data was constructed examining Bitcoin node for the years 2013 through 2016. The two main panel analyses (OLS and fixed effects) are listed below.

OLS:

$$nodes_{i,t} = \beta_0 + \beta_1 torRelay_{i,t} + \beta_2 elecPrice_{i,t} + \beta_3 internet_{i,t} + \beta_4 innovate_{i,t} + \beta_5 bankAccount_{i,t} + \beta_6 \log(pop_{i,t}) + \beta_7 \log(realGDP_{i,t}) + \beta_8 gdpGrowth_{i,t} + \varepsilon_{i,t}$$

Fixed Effects:

$$nodes_{i,t} = \beta_0 + \beta_1 torRelay_{i,t} + \beta_2 elecPrice_{i,t} + \beta_3 internet_{i,t} + \beta_4 innovate_{i,t} + \beta_5 bankAccount_{i,t} + \beta_6 \log(pop_{i,t}) + \beta_7 \log(realGDP_{i,t}) + \beta_8 gdpGrowth_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t}$$

Panel: Tor Relay

The first regression ran is an OLS on the panel data, and it yields results similar to the cross-sectional regressions. Tor relay has a positive and significant at the 0.1% level. This time, however, a 10-relay increase leads to a 0.001 increase in the number of Bitcoin nodes, which is noticeably smaller than the magnitudes derived previously. The innovation index is positive and significant at the 0.1% level, and population is again significant at the 1% level. The R-squared is slightly lower than in our previous results, with 28.93% of the variation in Bitcoin nodes across the period 2013-2016 being explained. (See Table 5, Regression 1)

In addition to the OLS regression, a Fixed Effects model is regressed, examining both country and time fixed effects. These results present an interesting twist – after country and time-fixed effects are taken into account, the sign of the coefficient on Tor relay switches. Now, a 10-relay increase corresponds to a .002 *decrease* in the number of Bitcoin nodes in a country, a value still statistically significant at the 0.1% level. The innovation index is significant at the 5% level. (See Table 5, Regression 2)

The sign flip of Tor relays brought about further research into why this is the case. In an attempt to identify an omitted variable (which is positively associated with both Tor relays and Bitcoin nodes), a second two-way fixed effects model is constructed. In this model, several new country-specific indices are included. Polity scores, developed by the Polity Project, measure a country's democratic standing and are included to examine whether government institutions can explain both nodes and relays. An economic freedom index is similarly included, with the reasoning that greater levels of economic freedom will correspond with more Bitcoin nodes and Tor relays. Lastly, both a corruption index and a money-laundering index are included – both of these indices are included as an extension of Athey et. al. (2016) and other literature, which identify illegal activity as a major proportion of Bitcoin and Tor use.

This model fails to adjust the negative coefficient on Tor relay. Neither democratic standing, economic freedoms, corruption, or laundering captured the negative effects. In this model, a 10-relay increase in Tor corresponds to a .003 decrease in the number of Bitcoin nodes in a country, a value statistically significant at the 0.1% level. The innovation index and real GDP are significant at the 5% level, and only the economic freedom index has a statistically significant effect, which was positive and significant at the 10% level. (See Table 5, Regression 3)

Extensive robustness checks attempting to explain the sign flip of Tor relays in the fixed effects models fail to successfully redistribute any significance or sign. When country fixed effects are included, the sign on Tor relay changes from positive to negative, while still retaining its significance at the 0.1% level. This suggests that the initial regressions in the paper are biasing the effect of Tor relays to be positive, akin to an omitted variable bias. In short, Tor relays are proxying for some unobserved, time-invariant, country-specific characteristic which is positively correlated to Bitcoin nodes and Tor relays. The robustness model (see Table 5, Regression 3) attempts to identity this characteristic through the inclusion of four additional indices. Admittedly, these indices are not time invariant. However, the model fails to identify the characteristic.

## DISCUSSION

### Limitations:

Using node distribution data for Bitcoin and Ethereum, this paper has shown that, contrary to popular belief, electricity price has not played a significant role in the distribution of nodes. However, three factors are limiting to this claim. First, data regarding electricity price was only available from Wikipedia, and although per the article each country estimate was obtained from official government sources, the nature of this source could have skewed the effects of electricity price upon node distribution in any number of directions. Most likely, the electricity price data for small or developing countries are misrepresented or inaccurate, however it is impossible to discern in which direction this misrepresentation may be biased. Second, the technologies examined in this paper are incredibly new, so historical trends are not definitive. The initial set of regressions examining electricity price are cross-sectional, and even the panel regressions examining the effect of Tor network relays upon node distribution, although more extensive in data than the cross-sectional regressions, still only examine the period 2013-2016. Extensive data on geographical node distribution, at this point in time, does not exist, so the actual effect of electricity price on node distribution should be examined later in time, perhaps in five to ten years, when historical trends and effects are observable.

Lastly, this paper recognizes that there is a possibility of reverse causality, in which Bitcoin nodes might explain Tor relays, instead of Tor relays explaining Bitcoin nodes, which would result in an endogeneity problem within the regressions conducted. The research in this paper relies on this endogeneity not existing, although it does not discount this potential. However, this paper has demonstrated a clear statistical relationship of Tor relays explaining the number of Bitcoin nodes, and in a historical sense, the Tor network has existed since the mid 1990s, and with Bitcoin being created in 2008, this reverse causality is very unlikely. In order to mathematically discount this potential endogeneity, we conducted a regression to identify an instrumental variable, *Z*, which would explain Tor relays but not Bitcoin nodes. A lack of time to pursue this inquiry yielded no significant conclusions. However, polity scores had a significant effect on the number of Tor relays, with a one-point increase in the polity score for a country corresponding to an 1885.71 increase in the number of Tor relays, a value statistically significant at the 1% level. As established in the previous regressions, polity scores did not have a significant effect on Bitcoin nodes, so further research into whether polity scores might function as *Z* for an instrumental variable regression is necessary. A summary of the regression explaining Tor relays as a function of the variables within this paper is listed in Table 6.

Discussion of Sign Flip:

This paper initially expected the electricity price to have a negative effect on node count, which was partly true, but only given certain internet level thresholds discussed previously. Furthermore, this research expected Tor relays to have a positive effect on node count, and although this effect was realized in the cross-sectional Tor relay models (and even the OLS panel data model), once country fixed effects were taken into account, the effect of Tor relay on

Bitcoin node count became negative. The OLS, time and country fixed effects, and time fixed effects – along with each regression's respective sign on Tor relays, are listed below:

OLS: Tor relay sign ( + )

$$Y_{i,t} = \alpha + \beta x_{i,t} + \varepsilon_{i,t}$$

Country and Time Fixed Effects: Tor relay sign ( - )

$$Y_{i,t} = \alpha + \beta x_{i,t} + \delta_t + \theta_i + \varepsilon_{i,t}$$

Time Fixed Effects: Tor relay sign ( + )

$$Y_{i,t} = \alpha + \beta x_{i,t} + \delta_t + \varepsilon_{i,t}$$

When country fixed effects, $\theta_i$, are taken into account, the sign of Tor relay changes to negative. Robustness checks attempting to reexamine this sign fail to yield different results, which suggests that the true effect of Tor relays on node count is in fact negative. Some unobserved, time-invariant, country-specific characteristic – which is positively related to both Tor relays and Bitcoin nodes – is being captured by the positive coefficient on Tor relays in the OLS and time-fixed effect panel models. Put another way, it is the variation in Tor relays *across* countries that produces the positive relation between the number of Tor relays and the number of Bitcoin nodes observed in the OLS and cross-sectional regressions. In contrast, it is the time variation *within* countries that produces the positive relation between the number of Tor relays and the number of Bitcoin nodes observed in the fixed effect regression. As the regression attempting to identify this characteristic with the additional indices failed to do so, and insufficient time inhibited further research into what this unobserved, time-invariant, country-specific characteristic might be, further research into this identification is necessary.

The mostly likely proposition for why Tor relays could have a negative effect on Bitcoin node count relates to the differences in the nature of secrecy in operating these technologies. People are very likely to be secretive of whether or not they are operating a Bitcoin node. Nodes are very expensive to create, and it is widely recommended not only to be secretive of operating a node, but even to be discrete about owning any cryptocurrency at all, due to the historical trends of cryptocurrency hacks. (Running, 2009) Tor relays, on the other hand, are likely to be much less discrete. The Tor Network details instructions on their website on how to operate a Tor relay, and explicitly advises those who do so to notify their internet service provider of the operation. (Legal FAQ, 2014) Because of the differences in privacy of Bitcoin nodes and Tor relays, we propose that individuals operating a Tor relay are less likely to operate a Bitcoin node. An individual might operate a Bitcoin node and use the Tor network, or operate a Tor relay and use Bitcoin for transactions, but due to the different natures of discretion with these technologies, an individual is unlikely to operate both a Tor relay and a Bitcoin node.

Broader Discussion:

The broader implications of this research are several. First, the fact that electricity price has not played a significant role up to this point in time is significant – it suggests that the distribution of these nodes is irrational, and Tor relays have contributed to this node distribution in a more significant manner. This suggests that mining is still an infant industry (Bitcoin was created only ten years ago), and the current distribution is akin to a "Gold

Rush" – individual agents, regardless of location or costs, are scrambling to set up mining operations due to the astronomical price increases, hoping to make a fortune. Over time, mining operations are likely to be consolidated. This will occur for two reasons: first, the current capital costs of setting up mining rigs are immense, as extreme demand increase for computer graphics cards (the electrical components used to create mining rigs) has caused a worldwide scarcity and associated price increases – more than doubling over the last year. (Martindale, 2018) Second, there will be diminishing marginal returns to mining Bitcoin as time goes on. This will be due to two causes: first, a continual growth in the number of Bitcoin users over time will continue to backlog network transactions, and cause the time to mine one coin to increase (and thus increase cost to mine one coin). Second, the rate at which new Bitcoins are created is programmed to issue less and less coins each year, issuing the very last Bitcoin around the year 2121. Large fixed costs, growing use, and slowing coin issuance will incentivize the consolidation of mining operations, where economies of scale make mining more profitable.

Once mining operations are consolidated and the "Gold Rush" nature of mining has ended, these consolidated operations will locate in places where electricity prices are low and internet availability is high. This paper projects Iceland to become the cryptocurrency mining center of the world for three reasons: first, electricity costs are minimal due to the sole use of geothermal and hydroelectric power. Second, Iceland has fast, widespread internet access, being the only country in this sample which has an internet access proportion of 100%. Lastly, along with the cost of electricity to mine, the cost to *cool* mining rigs is equally astronomical. A mining operation in Dubai, for example, might cost twice as much to operate compared to one in Iceland, with average summer temperatures of 96 and 53.5 degrees Fahrenheit, respectively. (Climate of Dubai, 2018) (Climate of Iceland, 2018) The consolidation of mining operations and the future significance of electricity price are two areas which require more research.

## CONCLUSION

This paper concludes that electricity price has not played a significant role in node distribution up to this point, but rather that Tor relays have played a much more significant role in this distribution. However, we also propose that the significance of electricity price in distribution is likely to increase each year, and further research into this increasing significance is necessary. As more data on node distribution around the world becomes available over time, the significance of electricity price might be manifested. Two specific areas of this paper which require more research are the identification of both the instrumental variable term, $Z$, and the unobserved, time-invariant, country specific characteristic responsible for the sign flip of the coefficient on Tor relays. Through this additional research, the potential endogeneity existing as a result of reverse causality will be discounted, and the characteristic which initiated the sign flip on Tor relays will be identified, providing a further understanding of the current distribution of Bitcoin and Ether nodes.

Cryptocurrencies have exploded in popularity and market capitalization within the last few years, and the electricity costs associated with this explosion have similarly skyrocketed. However, despite the massive electricity costs associated with mining, this paper has concluded that these costs have not played a significant role in Bitcoin and Ether node distribution up to this point. This research has provided the first academic look into the significance of electricity prices in mining operations, and has provided thorough analyses of the true drivers in Bitcoin and Ether node distribution around the world.

REFERENCES:

Akins, Benjamin W. and Chapman, Jennifer L. and Gordon, Jason M., A Whole New World: Income Tax Considerations of the Bitcoin Economy (November 7, 2013). Pittsburgh Tax Review, Forthcoming.

Athey, Susan, Iva Parashkevov, Vishnu Sarukkai, and Jing Xia. "Bitcoin Pricing, Adoption, and Usage: Theory and Evidence." Stanford Graduate School of Business. N.p., Aug. 2016. Web. 19 Sept. 2017.

Bajpai, Prableen. “Blockchain.” Investopedia, 18 Mar. 2018,

Balcilar, Mehmet. "Can Volume Predict Bitcoin Returns and Volatility? A Quantiles-based Approach." Economic Modelling. North-Holland, 27 Mar. 2017. Web. 19 Sept. 2017.

“Climate of Dubai.” Wikipedia, Wikimedia Foundation, 6 Apr. 2018.

“Climate of Iceland.” Wikipedia, Wikimedia Foundation, 19 Mar. 2018.

Committee on Payments and Market Infrastructures, comp. Digital Currencies. Rep. N.p.: Bank for International Settlements, 2015. Print.

“Definition of Cryptocurrency.” Oxford Dictionaries | English, Oxford Dictionaries, 2018.

Dyhrberg, Anne. "Bitcoin, Gold and the Dollar – A GARCH Volatility Analysis." Finance Research Letters. Elsevier, 11 Nov. 2015. Web. 19 Sept. 2017.

Hayes, Adam. "Cryptocurrency Value Formation: An Empirical Study Leading to a Cost of Production Model for Valuing Bitcoin." ResearchGate. N.p., May 2016. Web. 19 Sept. 2017.

Hern, Alex. “Bitcoin's Energy Usage Is Huge – We Can't Afford to Ignore It.” The Guardian, Guardian News and Media, 17 Jan. 2018

Hileman, Garrick, and Michel Rauchs. “Global Cryptocurrency Benchmarking Study.” Cambridge Center for Alternative Finance, 2017.

“Legal FAQ for Tor Relay Operators.” Tor Project, 21 Apr. 2014

Magee, Christine. "Bitcoin Blazes New Trail In Emerging Markets." TechCrunch. TechCrunch, 11 Feb. 2015. Web. 19 Sept. 2017.

Martindale, Jon. “Nvidia Speaks out against Rising Price of GPUs Due to Cryptocurrency Mining.” Digital Trends, Digital Trends, 22 Jan. 2018,

Nakamoto, Satoshi. “Bitcoin: A Peer-to-Peer Electronic Cash System.” Bitcoin.org, 2008. 25 Sept. 2017.

Pieters, Gina, and Sofia Vivanco. "Financial Regulations and Price Inconsistencies across Bitcoin Markets." Information Economics and Policy. North-Holland, 09 Feb. 2017. Web. 19 Sept. 2017.

Pilkington, Marc. (2016). Blockchain Technology: Principles and Applications.

Polasik, Michael & Piotrowska, Anna & Wisniewski, Tomasz & Kotkowski, Radossaw & Lightfoot, Geoff. (2014). Price Fluctuations and the Use of Bitcoin: An Empirical Inquiry. SSRN Electronic Journal.

“Running A Full Node.” Bitcoin Core, Bitcoin.org, 2009.

Yelowitz, Aaron & Wilson, Matthew. (2015). Characteristics of Bitcoin users: an analysis of Google search data. Applied Economics Letters. 22. 1-7.

Tables and Figures:

Table 3: Bitcoin Cross-Sectional

| **Bitcoin Cross-Sectional Regression** | 1 | 2 | 3 |
|---|---|---|---|
| Variables: | | | |
| elecPrice | 29.4<br>(32.41) | -1.21<br>(23.81) | -15.92<br>(26.54) |
| internet | 591.06<br>(1044.42) | -948.4<br>(787.7) | -951.6<br>(782.5) |
| innovate | 35.99**<br>(13.6) | 31.92***<br>(9.82) | 23.35**<br>(10.21) |
| log(gdpReal) | -119.22<br>(136.11) | -97.9<br>(98.2) | -135.8<br>(98.96) |
| gdpGrowth | -48.08<br>(37.29) | -43.38<br>(26.9) | -48.78*<br>(26.71) |
| log(pop) | 250.78*<br>(141.63) | 143.6<br>(103.4) | 220**<br>(107.3) |
| elecPrice x internet | -44.28<br>(43.39) | 11.17<br>(32.38) | 21.47<br>(33.69) |
| torRelay (10) | N/A | 0.036****<br>(0.056) | 0.037****<br>(0.005) |
| liberty | N/A | N/A | 196.1**<br>(83.93) |
| | | | |
| Observations | 55 | 55 | 55 |
| Adjusted R-squared | 0.3093 | 0.6409 | 0.6717 |
| Signif. Codes: *(0.1), **(0.05), ***(0.01), ****(0.001) | | | |

Table 4: Ether Cross-Sectional

| **Ether Cross-Sectional Regression** | 1 | 2 | 3 |
|---|---|---|---|
| Variables: | | | |
| elecPrice | 121.12 (81.75) | 45.78 (61.46) | -8.81 (68.53) |
| internet | 3075.62 (2634.76) | -712.7 (2033) | -1010 (2021) |
| innovate | 98.31*** (34.30) | 88.28*** (25.35) | 65.33** (26.36) |
| log(gdpReal) | -390.13 (343.38) | -337.7 (253.4) | -395.4 (255.5) |
| gdpGrowth | -123.34 (94.08) | -111.8 (69.43) | -395.4 (255.5) |
| log(pop) | 783.78** (357.29) | 520.1* (266.8) | 675.4** (277) |
| elecPrice x internet | -204.29* (109.47) | -67.84 (83.56) | -23.99 (86.99) |
| torRelay | N/A | .089**** (0.014) | 0.091**** (0.001) |
| liberty | N/A | N/A | 499** (216.7) |
| | | | |
| Observations | 55 | 55 | 55 |
| Adjusted R-squared | 0.354 | 0.6485 | 0.6786 |
| Signif. Codes: *(0.1), **(0.05), ***(0.01), ****(0.001) | | | |

Table 5: Panel

| **Panel Regressions** | | | |
|---|---|---|---|
| **Regression** | 1 | 2 | 3 |
| Variables: | | | |
| torRelay (10) | 0.001****<br>(0.0003) | -.0.003****<br>(0.0004) | -0.003****<br>(0.0004) |
| internet | -0.12<br>(0.12) | 0.23<br>(0.31) | 0.24<br>(0.34) |
| innovate | 0.9****<br>(0.23) | -1.46**<br>(0.65) | -1.75**<br>(0.74) |
| log(gdpReal) | -2.36<br>(2.49) | -61.59*<br>(31.76) | -85.76**<br>(35.95) |
| gdpGrowth | 0.16*<br>(0.42) | 0.47<br>(0.43) | 0.57<br>(0.46) |
| log(pop) | 6.42**<br>(2.63) | 88.0<br>(89.63) | 106.99<br>(95.61) |
| polity | N/A | N/A | -0.49<br>(1.10) |
| econFreedom | N/A | N/A | 1.74*<br>(1.01) |
| corruption | N/A | N/A | 0.34<br>(0.54) |
| launder | N/A | N/A | -3.19<br>(3.28) |
| | | | |
| Country Fixed Effects | No | Yes | Yes |
| Time Fixed Effects | No | Yes | Yes |
| Observations | 77 | 77 | 77 |
| Adjusted R-squared | 0.2893 | 0.2354* | 0.2548* |

Signif. Codes: *(0.1), **(0.05), ***(0.01), ****(0.001)
(*Regressions 2 & 3 show R-squared)

Table 6: Explaining Tor

| **Table 6: Explaining Tor** | |
|---|---|
| Regression | OLS |
| Variables: | |
| internet | 322.88<br>(302.58) |
| innovate | 1086.63*<br>(629.75) |
| log(gdpReal) | 5834.66<br>(6218.42) |
| gdpGrowth | -2243.59**<br>(991.09) |
| log(pop) | 11,203.49*<br>(6658.06) |
| polity | 1885.71***<br>(587.6) |
| econFreedom | 65.45<br>(513.67) |
| corruption | -382.24<br>(352.49) |
| launder | 4770.52<br>(3501.52) |
| | |
| Observations | 77 |
| Adjusted R-squared | 0.3237 |
| Signif. Codes: *(0.1), **(0.05), ***(0.01), ****(0.001) | |

Figure 1: Bitcoin Node Distribution

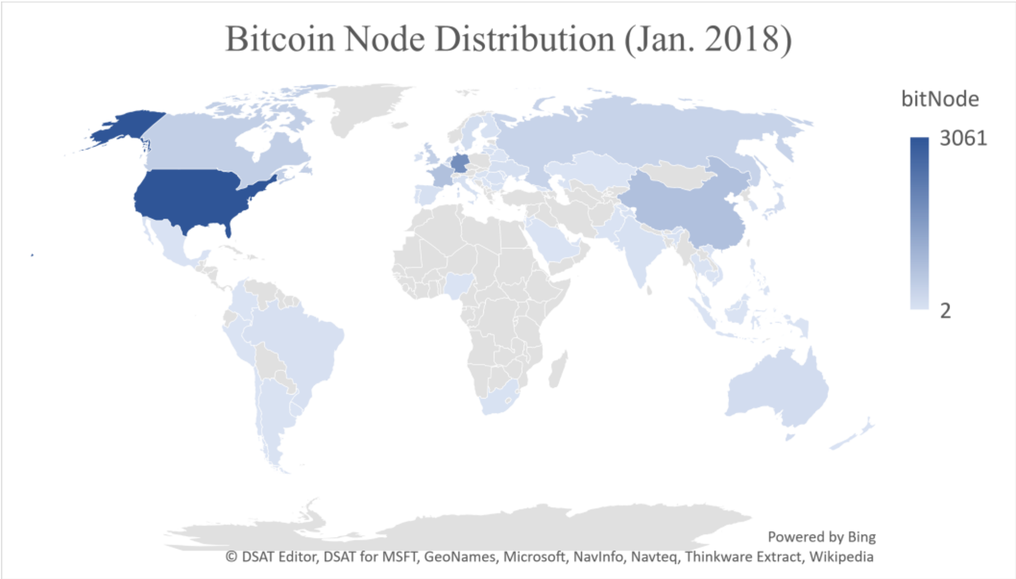


Figure 2: Ether Node Distribution

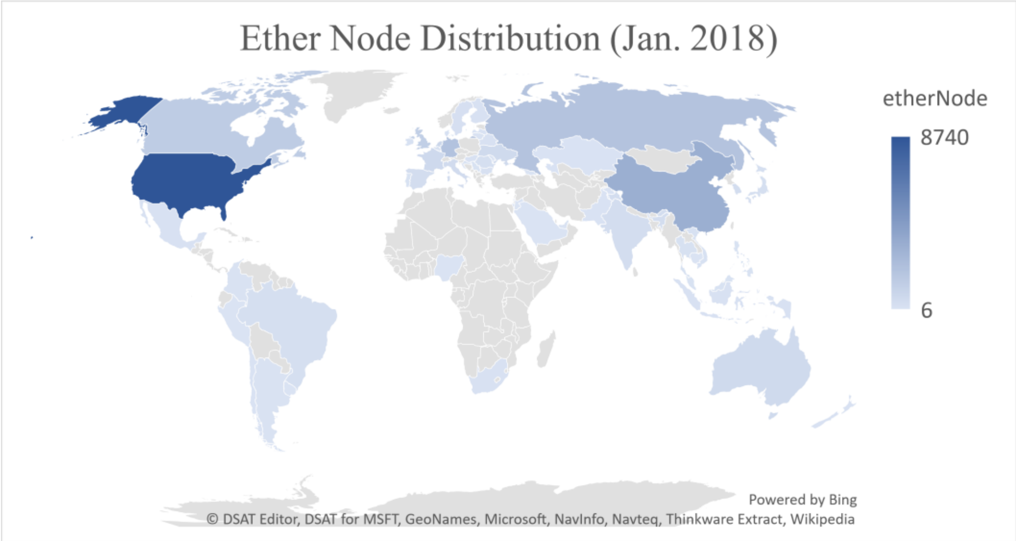

Figure 3: Tor Relay Distribution

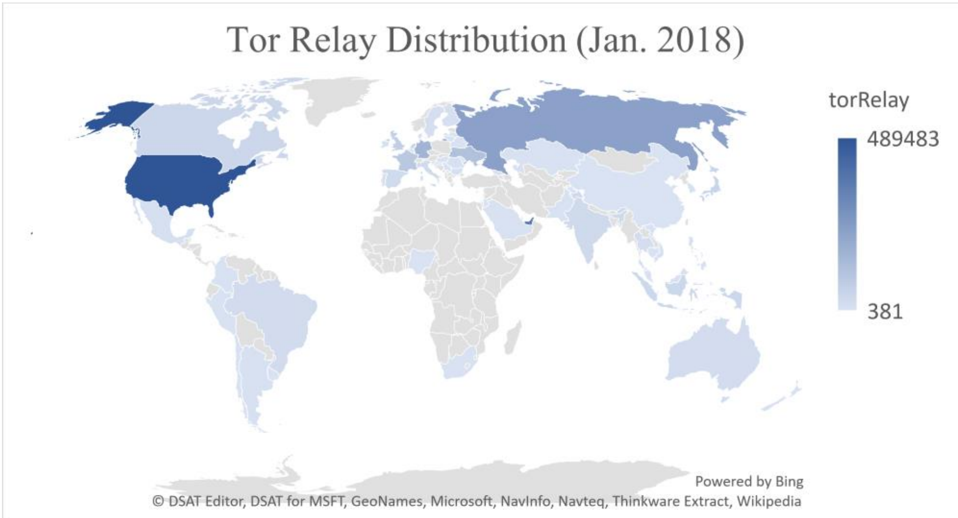


Figure 4: Electricity Price and Bitcoin Nodes

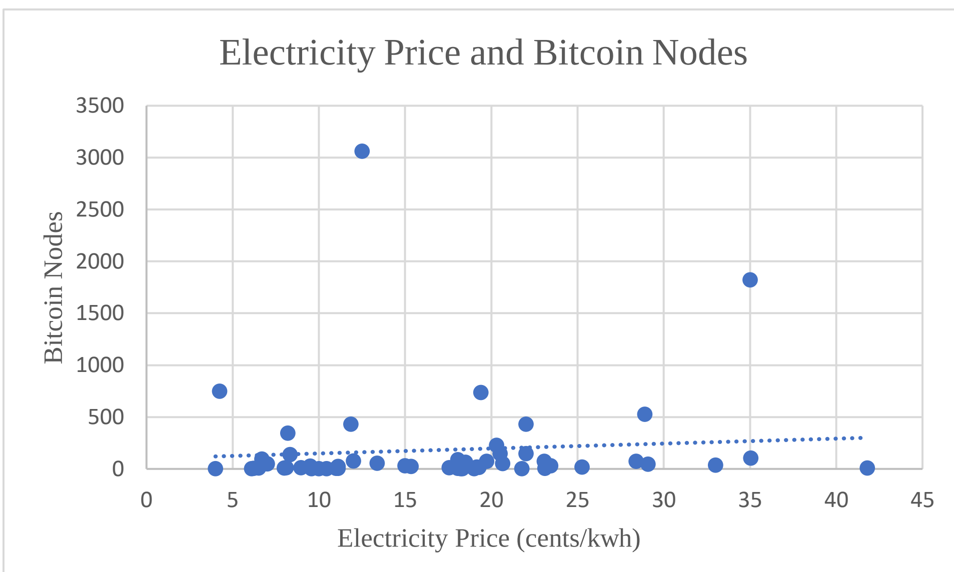

Figure 5: Electricity Price and Ether Nodes