**Replication and analysis of**

**“Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk”**

Huilan You

Institute for Quantitative Theory and Methods

Emory University

October 7, 2016

**Abstract**

 The original research paper, “Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk”, *Lecture Notes in Computer Science*, 2014 by Moore, Tyler; Christin, Nicolas examines factors that influence risks linked to Bitcoin exchanges, the platform to convert Bitcoins and hard currency. The researchers used best-fit Cox Model for survival analysis of exchange closure and logistic regression analysis for the exchange data breaches.

As a result, the paper suggests that an exchange’s average transaction volume is positively correlated with the probability of breach and is negatively correlated with the probability of its premature closure. By examining 40 Bitcoin exchanges, the paper concludes that high-volume exchanges are less likely to close but more likely to experience breach.

In my analysis, we replicate the original studies, utilize alternative methods on different parts of the analysis, and discover interesting discrepancies between our alternative analysis and the original analysis. With regard to the cox model of the original paper, we used logistic model instead and confirmed the result of the original cox model, demonstrating the robustness of the original analysis. we successfully discovered an influential point among the 40 exchanges and tried to condition on the location of the exchanges and test if the geographic location has influence on breach probabilities. Moreover, because we found out there is a correlation between operation months and daily volume, we suggest that the original paper should not have treated the two explanatory variables in the original logistic model as two independent variables.

**Introduction**  
 In the replication paper, our topic is the factors that affect Bitcoin-Exchange Risk. We are interested in this topic because of the increasing popularity of Bitcoin. Bitcoin became the most successful and popular decentralized digital currency and payment system worldwide as the alternative to hedge the risks from the banking system after the 2008 financial crisis. On one hand, Bitcoin’s market capitalization (Exhibit 1, Bitcoin.com) significantly increased from US$187 million in 2013 to US$121.2 billion. On the other hand, “Bitcoin has been repeatedly targeted by fraudsters” and many bitcoin trading platforms have experienced hacks and scams -“Since bitcoin’s creation in 2009 to March 2015, 33% of all bitcoin exchanges operational during that period were hacked funded by the U.S. Department of Homeland Security” (Fortune). Therefore, we selected the original paper, “Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk” for our analysis as it provides readers insights about closure and data breach of the Bitcoin exchanges. The goal of our project is to meticulously examine the results of the original research and discover interesting findings that the authors of the original paper missed by utilizing alternative rigorous methods and approaches. By doing so, we hope to give the readers, researcher and investors the comprehensive perspectives of the current status of the Bitcoin exchanges.

**Replication**

The main purpose of the original research, “Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk” (Moore and Christin, 2014), is to examine the factors that affect the closure and the data breach of Bitcoin exchanges. Bitcoin exchanges are platforms for investors to exchange between Bitcoins and hard currency. The original paper shows that average daily transaction volume is negatively correlated with the survival time and probability of closure of an exchange. In addition, it also provides analysis indicating that exchanges with higher transaction volume are more likely to suffer security breach. Using the same models and methodology, we successfully replicate the results from the original paper using the provided dataset and R code.

Our analysis is based on the same dataset of the 40 Bitcoin exchanges which the original paper used. The Bitcoin exchange rates in the original research paper is collected from the website*, bitcoincharts.com*. The average daily transaction volume and the lifetime of an exchange are calculated as *log(daily vol)*, because the original paper analyzes the relationship between the percentage change of the daily transaction volume and the dependent variables. If the exchange is closed by January 16, 2013, then it is considered as closed. In addition, the original paper also collects reports from Bitcoin forums to check if an exchange has experienced security breach and if the exchange repaid its customers after the breach. Lastly, the original paper collects the “Anti-Money-Laundering and Combating the Financing of Terrorism” (AML/CFT compliance) that is computed by World Bank economists. AML/CFT compliance is an important factor because it describes a country’s anti-money-laundering policy, which may affect the operation of the Bitcoin exchange.

In the first part, the original research constructs a survival analysis that examines the influence of transaction volume, occurrence of security breach, and AML/CFT compliance on the lifetime of an exchange. The best-fit Cox model describes how the survival probability (dependent variable) of an exchange changes over time. The model suggests that daily volume is negatively associated with the hazard rate. If an exchange doubles its daily transaction volume, it will have 16% reduction in the hazard rate. The other two factors: the occurrence of security breach and AML/CFT compliance, are not significantly correlated with the survival probability. The dataset shows that among the exchanges that experienced a security breach, about 55% of them closed. The percentage also provides little indication that the closure of an exchange is correlated with the occurrence of breach. The p-value of AML/CFT is 0.9221, which suggests that the anti-money laundering policy of a country does not have significant influence on the hazard rate.

The second model examines the factors that influence the possibility of the security breach. The author chose logistic model because the dependent variable(whether data breach occurs) is binary and categorical and there are two quantitative explanatory variables in this model(daily volume and months operational). The result suggests that the daily volume is positively correlated with possibility of breach (p=0.0131, coefficient=0.514) while correlation between months operational and possibility of breach is less significant (p=0.14, coefficient=-0.104).

**Analysis**

***Proportional hazard model***

**I. Alternative logistic model for the original Cox model**

Average daily transaction volume determines if an exchange is profitable. Exchanges with lower average daily transaction volume make less profits, which make them less sustainable and easier to shut down. The second key factor considered is if an exchange has experienced a security breach. A security breach may exert significant influence on the stability of a Bitcoin exchange. For example, if an exchange experiences a security breach, the exchange will suffer the loss of profits and cash flow. Customers will ask the exchange for reimbursement, which further inflicts an exchange’s financial condition. In addition, the exchange will also lose its credits and prospective customers. Lastly, different countries have different AML/CFT compliance, which describes a country’s emphasis on anti-money laundering policies. Due to Bitcoin’s intractableness, bitcoin is a popular tool for money laundering. If a Bitcoin exchange is in countries that are more restrictive on money-laundering activities, it may experience greater restriction and suppression from the government. As a result, a country’s AML/CFT compliance may also influence the survival time of a Bitcoin exchange. In our analysis, we try to examine the relationship between the closure of an exchange and the above-mentioned variables using different models to check the robustness of the original findings. In the first alternative model, we try to find if transaction volume, occurrence of security breach, and AML/CFT compliance have an influence on the probability of the closure.

The original survival model is a proportional hazard model that describes the change of the risk over time. Therefore, we try to check if a linear regression of the lifetime of an exchange on transaction volume, occurrence of security breach, and AML/CFT compliance also provides similar results.

The first model takes the following form:

The linear regression yields the following results:

Table 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Intercept | log2(daily transaction vol) | Security breach | AML/CFT compliance score |
| Coefficient | 324.199 | 31.257 | -111.323 | -5.530 |
| Standard Error | 177.48 | 9.771 | 86.603 | 6.242 |
| Significant | 0.07598 | 0.00288 | 0.20685 | 0.3816 |

The original Cox-Model suggests that an increase in daily transaction volume will decrease the hazard rate of an exchange. Our alternative linear regression model suggests that an increase in daily transaction volume will increase the lifetime of an exchange. Both occurrence of breach and AML/CFT compliance are not significantly correlated with the lifetime of an exchange. Since a decrease in hazard rate can potentially lead to longer lifetime of an exchange, our model supports the finding of the original paper and provides an alternative perspective to look at the closure of an exchange.

In addition, we are interested about if the transaction volume is also correlated with the probability of the closure. The closure of an exchange is a binary outcome, so we use a logistic model to test if the transaction volume an influence on the probability of an exchange’s closure.

The model takes the following form, where p is the probability of an exchange’s closure:

Logistic regression yields the following results:

Table 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Intercept | Log2(daily transaction vol) | Security breach | AML/CFT compliance score |
| Coefficients | 0.878 | -0.0306 | 0.265 | -0.00959 |
| Standard Error | 0.428 | 0.0236 | 0.209 | 0.0151 |
| Significance | 0.0475 | 0.201 | 0.21 | 0.528 |

None of the variables are significantly correlated with the probability the closure. The result suggests that transaction volume may only have an influence on the lifetime of an exchange but cannot tell if an exchange will close.

***Logistic Model for the probability of breach***

1. **Variable Analysis**

The second model examines the factors that influence the possibility of breach. The author suggests that the possibility of breach is positively correlated with transaction volume and how many months that an exchange operates. Since higher transaction volume implies that an exchange makes more profits, higher transaction volume makes an exchange a more lucrative target for hackers and malignant attacks. Thus, exchanges with higher transaction volume are more likely to experience a security breach. Secondly, the original research considers how long an exchange operates. It is intuitive to think that the longer an exchange remains operational, the more probable an exchange will experience a security breach.

The logistic model in the original paper only focus on two explanatory variables that would influence the dependent variable, whether a breach occurs. However, our research indicated that there are other explanatory variables the original paper omitted. For instance, “the ability of the IT team for different exchanges” can be a forgotten explanatory variable. While “the ability of the IT team” intuitively has a week or even zero correlation with “daily volume”. As “daily volume” tends to be more related to the current financial market and investors’ behavior, it is highly possible that the ability of the IT team is independent to the explanatory variables the original paper considered. In addition, since 56% of surveyed small businesses was reported to experience cyber-attacks, “a successful IT team should be knowledgeable of modern cybersecurity techniques and programs” (Beck, TECHCO). Therefore, “the ability of the IT team” affects the probability of a breach occurs and may be a potential explanatory variable the original paper omitted. In real life, there may be other explanatory variables the original paper ignored. Although we were unable to quantify and include every variable into our analysis, we admitted that the result of the original paper and the result of our analysis could not be perfectly rigorous due to the risk of ignoring these forgotten explanatory variables.

Moreover, it is also possible that there are omitted variables which influence both the independent variables (daily volume or months operational) and the dependent variable (whether a breach occurs). In fact, there are many potential confounding variables could be hold in this context. For example, the number of reports about Bitcoin-Exchange could have correlations with both independent variable (average daily transaction volume) and dependent variable (probability of breach). The confounding variable (reports) may be correlated with the average daily transaction volume because more and more people are going to invest their money in Bitcoin after reading the reports and articles. The number of reports can also draw hackers’ attention and increases the probability of being hacked. Hence, such omitted variable has the risk to lead our statistical analysis to become biased. Although we cannot control all the omitted variables in our model. We carefully tested on the two explanatory variables (daily volume and months operational) the original paper suggested and designed a new logistics regression model to control the uncertainty on location.

First, we analyzed the relationship between the two explanatory variables(daily volume and months operational) and the relationship between the secondary variable(months operational) and the dependent variable(whether breach occurs). The reason we did these two analyze was that while the original paper treats “months operational” as a secondary explanatory variable, we recognized that it is possible that “months operational” is actually an omitted variable if it correlated to both “daily volume” and “whether breach occurs”.

In order to examine the relationship between the original dependent variable and “months operational”, we ran the logistic model so that “whether breach occurs” solely regressed on “months operational”. The coefficient of “months operational” is-.004636 with extremely high P-value, 0.927. Thus, the relationship is not significant and the analysis suggests that there is negatively weak relationship between “months operational” and “whether breach occurs”. Comparing the the logistic analysis the original paper conducted, we discovered that the original states that “months operational is negatively correlated with being breached(coefficient=-.104), but the association fall short of statistical significance(p=0.14) ”. Even though there is a discrepancy between these two coefficients, this discrepancy may be mainly caused by the difference between univariate logistic model and multivariate logistic model. More importantly, both original paper and our analysis suggests that there is no statistically significant relationship between “month operational” and “whether breach occurs”

Then, in order to examine the relationship between “months operational” and “daily volume”, we utilized the function cor in R. Since the original paper descended the power of “daily volume” by log, we kept the consistency and found the correlation between log “daily volume” and “months operational”. As a result, the correlation is moderate (0.4095), which implied that there is a moderate dependence between “month operational” and log “daily volume.” Therefore, the original paper may have statistical errors since it considered these two variables as two independent explanatory variables.

**II. Influential point**

The original logistic model suggests that daily transaction volume is positively associated with the probability of the occurrence of breach. However, it is important to first examine the daily transaction volume dataset. At first glance, there are several extreme values. While the exchange Vircurex has daily transaction volume as low as 6, Mt. Gox has a high daily transaction volume of 43230. We believe that these extreme points may affect the original logistic model. One way to test the influence of an outlier is to compute the regression equation with and without the outlier.

The original paper shows that if an exchange doubles its daily transaction volume, its log odds of occurrence of security breach will increase by 0.514 (p=0.0176). We are wondering if there is any specific bitcoin exchange that greatly affects this result. In another words, we need to find out if any of bitcoin exchange stands as an outlier and thus influences the coefficient of the daily transaction volume variable. We reconstructed the model by excluding one exchange each time for 40 times and observe if the logistic model change significantly.

Our analysis shows that if we run the logistic model excluding the exchange Vircurex, the coefficient of the daily transaction volume variable changes greatly. While all the other coefficients range from 0.46 to 0.54, the logistic model without the exchange Vircurex has a coefficient of 1.6199 (p=0.0283). This suggests that if an exchange doubles its daily transaction volume, the log odds of the occurrence of security breach increases by 1.6199, which is significantly different from the original finding (0.514). The outlier that we found in this analysis is Vircurex. Vircurex has a low daily transaction volume of 6 but has also experienced a security breach, which is contradicted to the result the original paper suggests (daily volume is positively correlated to the probability of data breach)

In addition, in the original paper, the logistic model suggests that the number of operational month is not significantly correlated with the possibility of security breach(p=0.14). However, when we do not include the exchange Vircurex, our logistic model suggests that the number of operational month is significantly correlated with the possibility of security breach (p=0.0423). The coefficient of the number of operational month suggests that if an exchange increases its operational month by 1 unit, its log odds of the occurrence of security breach decreases by 0.3123. This is an interesting observation because the original paper hypothesizes that the longer an exchange operates, the more probable it will suffer a security breach. However, our new logistic model implies that the longer an exchange operates, the less probable it will suffer a security breach. One explanation for this result is that exchanges that operate longer have more experience and better security system, which make an exchange more immune to security attacks.

**Limitation**

Because we only have 40 bitcoin exchange data as same as the original paper, such small amount of data led our analysis to have a large standard error, which means a higher variability of our dataset. Moreover, the small sample may be not representative of all the Bitcoin exchanges in the world and as a result, our analysis may be biased. In addition, because maximum likelihood estimates are less robust than the ordinary least squares (linear regression), as a maximum likelihood estimate, logistic regression needs at least 10 cases of each independent variable, “some statisticians recommend at least 30 cases for each parameter to be estimated (Statistics Solutions)”. Having only 40 bitcoin exchanges decreases the reliability our result when we tried to run model condition on confounding variables.

**Conclusion**

Our analysis examined the correctness and robustness of the research paper “Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk”. The first model in the paper examines the factors that influence the hazard rate of an exchange’s closure and survival time. We firstly used two alternative models to examine the robustness of this model. A linear regression of an exchange’s lifetime on its daily transaction volume, occurrence of security breach, and AML/CFT compliance suggests that an exchange’s survival time is significantly correlated with its daily transaction volume. This model confirms and supports the finding in the original Cox model. The second alternative model is a logistic regression of the probability of an exchange’s closure on its daily transaction volume, occurrence of security breach, and AML/CFT compliance. This model returns no significant results and suggests that transaction volume may only have influence on the lifetime of an exchange but cannot tell if an exchange will close.From our analysis of the second model, we discovered that 1) an outlier(Vircurex) among the 40 Bitcoin exchanges that is contradictory to the original result, 2) there is a correlation between operation months and daily volume and 3) geographic location does not have significant effect on the probability of data breach by running.

**Appendix**

Exhibit 1

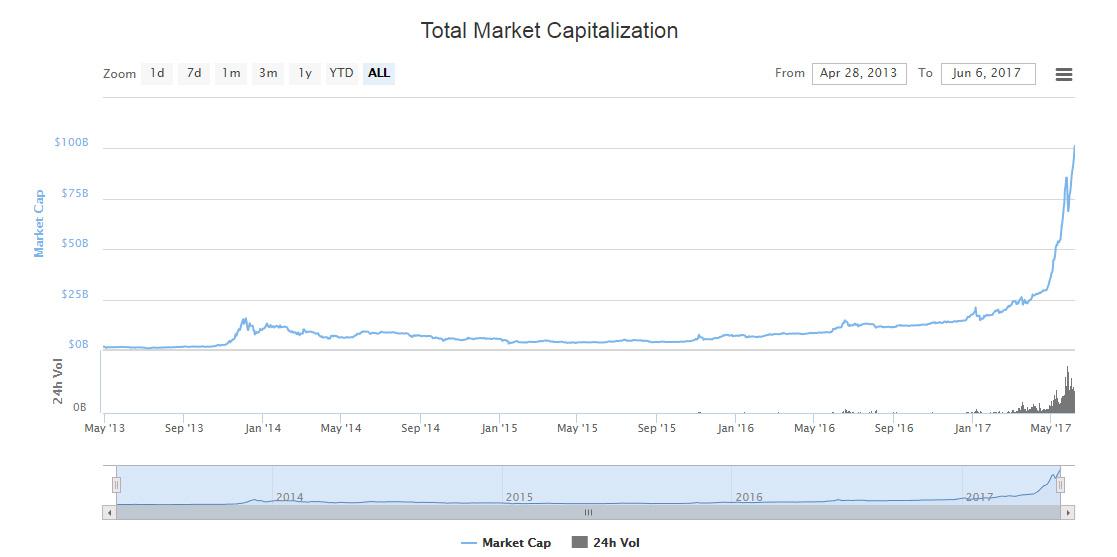
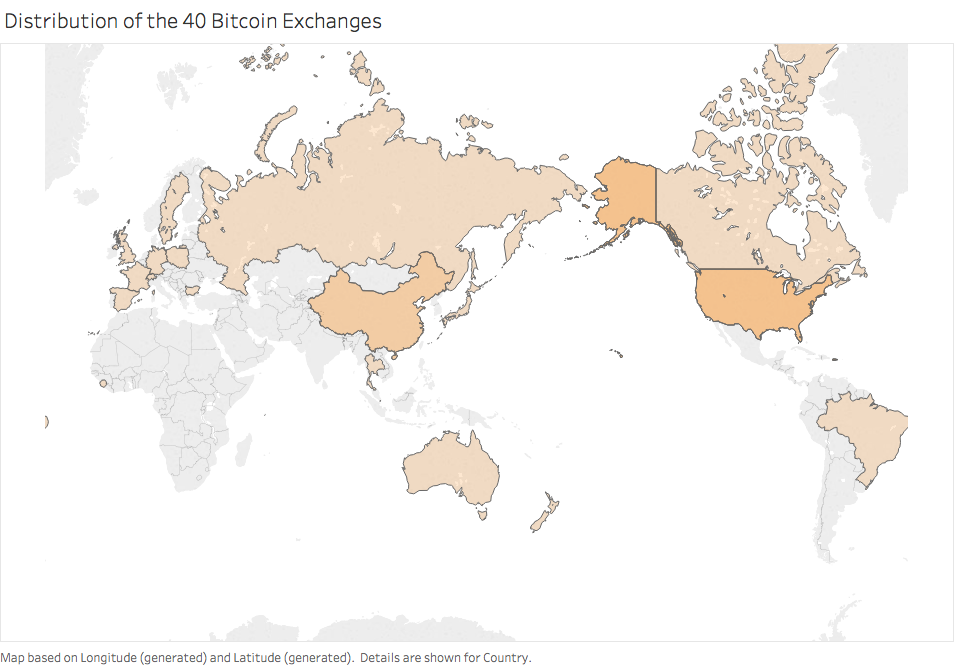


Exhibit 2



Original Dataset:

Exchange Origin Dates Active Daily vol. Closed? Breache? Repaid? AML Risk Ratio

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BitcoinMarket | US | 4/10 – 6/11 | 2454 | yes | yes | – | 34.3 | 1.12 |
| Bitomat | PL | 4/11 – 8/11 | 758 | yes | yes | yes | 21.7 | 1.28 |
| FreshBTC | PL | 8/11 – 9/11 | 3 | yes | no | – | 21.7 | 2.01 |
| Bitcoin7 | US/BG | 6/11 – 10/11 | 528 | yes | yes | no | 33.3 | 1.59 |
| ExchangeBitCoins.com | US | 6/11 – 10/11 | 551 | yes | no | – | 34.3 | 0.65 |
| Bitchange.pl | PL | 8/11 – 10/11 | 380 | yes | no | – | 21.7 | 0.61 |
| Brasil Bitcoin Market | BR | 9/11 – 11/11 | 0 | yes | no | – | 24.3 | 3.85 |
| Aqoin | ES | 9/11 – 11/11 | 11 | yes | no | – | 30.7 | 1.57 |
| Global Bitcoin Exchange | ? | 9/11 – 1/12 | 14 | yes | no | – | 27.9 | 1.45 |
| Bitcoin2Cash | US | 4/11 - 1/12 | 18 | yes | no | – | 34.3 | 1.47 |
| TradeHill | US | 6/11 - 2/12 | 5082 | yes | yes | yes | 34.3 | 0.94 |
| World Bitcoin Exchange | AU | 8/11 – 2/12 | 220 | yes | yes | no | 25.7 | 1.80 |
| Ruxum | US | 6/11 – 4/12 | 37 | yes | no | yes | 34.3 | 1.24 |
| btctree | US/CN | 5/12 – 7/12 | 75 | yes | no | yes | 29.2 | 0.98 |
| btcex.com | RU | 9/10 – 7/12 | 528 | yes | no | no | 27.7 | 0.61 |
| IMCEX.com | SC | 7/11 – 10/12 | 2 | yes | no | – | 11.9 | 1.88 |
| Crypto X Change | AU | 11/11 – 11/12 | 874 | yes | no | – | 25.7 | 0.53 |
| Bitmarket.eu | PL | 4/11 – 12/12 | 33 | yes | no | no | 21.7 | 1.09 |
| bitNZ | NZ | 9/11 – pres. | 27 | no | no | – | 21.3 | 1.14 |
| ICBIT Stock Exchange | SE | 3/12 – pres. | 3 | no | no | – | 27.0 | 2.15 |
| WeExchange | US/AU | 10/11 – pres. | 2 | no | no | – | 30.0 | 2.23 |
| Vircurex | US? | 12/11 – pres. | 6 | no | yes | – | 27.9 | 4.41 |
| btc-e.com | BG | 8/11 – pres. | 2604 | no | yes | yes | 32.3 | 1.08 |
| Mercado Bitcoin | BR | 7/11 – pres. | 67 | no | no | – | 24.3 | 0.95 |
| Canadian Virtual Exchange | CA | 6/11 – pres. | 832 | no | no | – | 25.0 | 0.53 |
| btcchina.com | CN | 6/11 – pres. | 473 | no | no | – | 24.0 | 0.60 |
| bitcoin-24.com | DE | 5/12 – pres. | 924 | no | no | – | 26.0 | 0.52 |
| VirWox | DE | 4/11 – pres. | 1668 | no | no | – | 26.0 | 0.45 |
| Bitcoin.de | DE | 8/11 – pres. | 1204 | no | no | – | 26.0 | 0.49 |
| Bitcoin Central | FR | 1/11 – pres. | 118 | no | no | – | 31.7 | 0.91 |
| Mt. Gox | JP | 7/10 – pres. | 43230 | no | yes | yes | 22.7 | 0.49 |
| Bitcurex | PL | 7/12 – pres. | 157 | no | no | – | 21.7 | 0.76 |
| Kapiton | SE | 4/12 – pres. | 160 | no | no | – | 27.0 | 0.80 |
| bitstamp | SL | 9/11 – pres. | 1274 | no | no | – | 35.3 | 0.54 |
| InterSango | UK | 7/11 – pres. | 2741 | no | no | – | 35.3 | 0.45 |
| Bitfloor | US | 5/12 – pres. | 816 | no | yes | no | 34.3 | 1.45 |
| Camp BX | US | 7/11 – pres. | 622 | no | no | – | 34.3 | 0.63 |
| The Rock Trading Company | US | 6/11 – pres. | 52 | no | no | – | 34.3 | 1.14 |
| bitme | US | 7/12 – pres. | 77 | no | no | – | 34.3 | 1.04 |
| FYB-SG | SG | 1/13 – pres. | 3 | no | no | – | 33.7 | 2.23 |

**Reference**

Reuters. “Risk of Bitcoin Hacks and Losses Is Very Real.” Fortune, 29 Aug. 2016, fortune.com/2016/08/29/risk-of-bitcoin-hacking-is-real/.

Althauser, Joshua. “Chicago Board Options Exchange to Launch Bitcoin Futures, Announces Specifics.” *Cointelegraph*, Cointelegraph, 25 Nov. 2017, cointelegraph.com/news/chicago-board-options-exchange-to-launch-bitcoin-futures-announces-specifics.

Beck, Christopher. “The IT Team Is Your Company's Secret Weapon.” *TechCo*, 18 Jan. 2015, tech.co/security-team-secret-weapon-2015-01.