Existence of online Ponzi schemes: Insights into new measurements

Sharad Agarwal

Supervisor: Dr Marie Vasek

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Master of Science in **Information Security**
at
University College London.

Department of Computer Science
University College London

Academic Year 2020/2021

I, Sharad Agarwal, confirm that the work presented in this thesis is substantially the result of my own work except where explicitly indicated in the text. The report may be freely copied and distributed provided the source is explicitly acknowledged.

Abstract

Ponzi schemes have been running since the mid-1800s. Since most people started using the internet widely, almost everything shifted online, including investment frauds - Ponzi schemes. Previous research has shown that Ponzi schemes form a significant majority of online financial frauds. This report focuses on understanding online Ponzi schemes/High Yield Investment Programs (HYIPs) continuous presence and presents various possible reasons behind their existence in today's world. One of the new insights is that 62.8% of all collected HYIPs claim to be in the United Kingdom (UK). 56% of all investigated HYIPs are officially registered in the UK as a 'limited company' with a registration number provided by the UK Companies House. Even though the UK Financial Conduct Authority (FCA) is the one that regulates the companies conducting finance-related activities, the UK Companies House provides a certificate of registration to a company without any document verification. The report provides a geographical map plotting the addresses of the HYIPs claiming to be in the UK. We also check their validity and similarities as we collect these addresses and use them for further analysis. We further study the collected data, including the HYIPs' social media platforms and payment processors. The lifetime of the HYIPs helps to understand the success/failure of the investment schemes. It also points out which scheme was able to attract more victims/investors. The analysis reveals that having a valid UK address, accepting currencies like Ethereum, Litecoin, and Perfect Money, and using the Twitter social media platform affect HYIPs' lifetime.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr Marie Vasek, for her continuous support, guidance, and recommendations. I would not have been able to complete this Master's thesis without it.

I would also like to thank my family and friends for their love and support, especially during this challenging time of the COVID'19 pandemic, motivating me every time and bringing out the best in me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The internet provides an accessible medium to people across the globe to easily invest in various online investment schemes. The widespread use of the internet has shifted many businesses online, including the classic Ponzi schemes to online Ponzi schemes/ High Yield Investment Programs (HYIPs). With this shift and popularity of HYIPs, a substantial amount of money in several currencies, including cryptocurrencies and other digital currencies, has been lost by investors. Being one of the significant online financial frauds, it has affected thousands of people worldwide [1]. Ponzi schemes cause damage to the economy, and most countries prohibit them by creating strict laws. Even though quite a lot of research has been done in this area and payment processors like Liberty Reserve have been shut down [2], the fraud continues.

Researchers have collected data by scraping various online discussion forums. Many machine learning algorithms have been experimented with to detect HYIPs running on Bitcoin and Ethereum blockchains. The discovery of various factors like trust in referees and scammer interactions shows why users invest in these schemes. However, loopholes are being exploited that remain undiscovered. Therefore, even after conducting so much research providing valuable insights, one crucial question unanswered is why many HYIPs still exist and succeed in scamming users? Something is missing in the research landscape, which the scammers are taking advantage of and earning money. Understanding the lifetime of the HYIPs is vital to see which schemes are successful in luring more victims. We are interested in

finding new factors that influence the lifetime of the HYIPs, making them exist in today's era.

Registration of a company forms an essential factor to have trust in a company. People often check whether the company there are investing in is legitimate or not. This leads us to think about more questions - Do these online Ponzi schemes register themselves as an official company? How do they manage to register themselves as a company to operate online? Which countries do scammers primarily choose for running such schemes? Does social media or using a particular payment factor influences the HYIP's lifetime? Based on the literature review [3], we have done in-depth research over the past few months and provide answers to these questions by diving deeper into the HYIP schemes currently running on the internet.

1.1 Contributions

After collecting an enormous amount of data, sanitizing and analyzing it, we present the following contributions in this report:

- We collected data about 366 High Yield Investment Programs (HYIPs) by crawling the aggregators - *hyip.com* and *hyiprank.com*. We identified and created a list of countries where the HYIPs claim to be based. 62.8% of these HYIPs claim to be in the United Kingdom.

- We build an updated HYIP dataset that includes new features like their address, company registration number, contact details, social media handles, and various currencies they accept. We also check if the HYIP is licensed by one of the most common HYIP kit developers - Goldcoders. We summarise these variables in Section 4.2.

- We understand and visualize the survival function for the lifetime of the HYIPs, helping to recognize HYIPs' success and failure. The survival analysis and Cox proportional hazards model in Section 5 shows that a valid UK address, accepting digital currencies like Litecoin, Ethereum, and Perfect Money, using a Twitter handle, and the number of payment processors accepted by HYIPs affects the lifetime of HYIPs.

- We verify and provide the numbers of HYIPs which are officially registered as a limited company in the countries they claim to be from. 56.01% of the total HYIPs are confirmed to be registered in the United Kingdom as a limited company and have a certificate of registration provided by the UK Companies House. Other countries where the HYIPs are registered include the Republic of Seychelles, the Republic of Marshall Islands, Australia, United States of America, Hong Kong, Belize, British Virgin Islands, and Saint Vincent and the Grenadines.

- We also identify in the verification process that some companies use a photoshopped Companies House certificate of registration pretending to be an officially registered limited company. It is a new finding as it points out that scammers want to gain investors' trust by showcasing that they are an officially registered company.

1.2 Organization of the report

We question the existence of HYIPs in the current era. This report aims to discuss and explore the findings of why online Ponzi schemes/HYIPs' still exist and what factors affect HYIPs' lifetimes. It is divided into six sections. Section 1 provides an introduction to this subject and lists the contributions of this report. Section 2 provides an understanding of Ponzi schemes and HYIPs with some examples of the famous Ponzi schemes. The related work done in this area with their significant contributions and critical analysis has been highlighted in Section 3. Section 4 explains the data collection methodology we use and summarises the collected data. We provide a statistical analysis of the data in Section 5. Finally, Section 6 concludes the report and provides recommendations for future work.

Chapter 2

Background

Since the 1800s, both legitimate investment schemes and investment frauds have existed. People these days want to invest money but do not necessarily understand what investments are secure. With so many forms of available investment opportunities comes great responsibility. The investment opportunities presented by such scams may consist of equity stakes, yachts, various forms of currencies (crypto and foreign), or debt issued by suspicious companies. These are often purportedly backed by a new product like a new cryptocurrency, technology, medical research, or business opportunity [4]. One such fraudulent investment scheme is called the Ponzi scheme. Scammers lure people into investing money in such schemes by promising them very high returns, cautioning a minimal risk. Ponzi schemes mainly depend on the new investors joining the scheme to pay the desired profits to the already joined investors. With the rise of technology and the internet, these fraudulent schemes have shifted online and are often called High Yield Investment Programs (HYIPs). These schemes collapse as soon as there are no new investors as there is not enough money to pay out the returns.

2.1 Ponzi scheme

Ponzi scheme takes advantage of people who invest money by luring the investors into their scheme and convincing them to pay high profits. It generally requires an initial investment and promises more than the average rates of returns. The U.S. Securities and Exchange Commission states that most Ponzi schemes focus on at-

tracting new investors to have an inflow of money to make promised payments to earlier investors and divert some of these invested funds for personal use instead of engaging in any legitimate investment business [5]. Almost with no legitimate income, Ponzi schemes require a consistent in-flow of money to survive. When it becomes challenging to achieve more investors or enormous numbers of existing investors withdraw, these schemes tend to collapse [6]. A wide variety of new businesses, investment vehicles, loans, and strategies typically legitimate have become the basis of Ponzi schemes.

2.1.1 Charles Ponzi

Ponzi scheme scams have been running for more than a century and were termed 'Ponzi schemes' in the 1920s on the name of a swindler called Charles Ponzi. An international reply coupon (IRC) was a coupon that could be exchanged for several priority airmail postage stamps from another country. Ponzi received a letter in the post from a company in Spain that enclosed this IRC. By purchasing IRCs in one country and trading them for more valuable stamps in another country as a form of arbitrage, he realized that he could make an enormous profit [7]. Ponzi figured that by finding a way to trade with the coupons in a large quantity, one could become rich [8]. He started transferring money to representatives working for him in other countries, buying the IRCs and shipping them back to the United States. Ponzi reportedly made more than 400 percent on some of these sales [7]. Ponzi convinced a few investors to fund his start-up money, assuring them a 50% profit in 45 days or 100% in 90 days. This marked the beginning of the scheme that carries Ponzi's name to this day [8]. Ponzi promised investors' returns for what he claimed was an investment. Ponzi used these funds from the new investors to pay fake returns to previous investors [9].

The scheme lasted until August of 1920 when The Boston Post began investigating Ponzi's Securities Exchange Company [10]. As a result of the newspaper's investigation, Ponzi was arrested by federal authorities on August 12, 1920, and charged with several counts of mail fraud [7].

2.1.2 MMM Ponzi scheme

Another one of the world's biggest Ponzi schemes is MMM. The MMM cooperative was started by Sergei Mavrodi, his brother Vyacheslav Mavrodi, and Olga Melnikova in early 1990 [11]. The scheme ran for a few years and acquired 15 million investors across the country in the meantime [12]. Shares and coupons from MMM were circulated as a parallel currency to the ruble and foreign currencies. They were even exchanged for food and clothes. In 1994, the company was accused of tax evasion, leading it to collapse. However, Mavrodi brought the scheme into existence in the 21st century (around 2011) as a High Yield Investment Program. The scheme was getting an in-flow of millions of dollars per day at its peak. It also became prevalent in Nigeria in early 2017. The scheme was marketed in Asia and Africa as a fund for mutual aid to help people selflessly, promising 30% monthly returns [11]. MMM is a classic example of how Ponzi schemes start, collapse, and come back after a few years.

2.1.3 Bernard L. Madoff Ponzi Scheme

History's largest Ponzi scheme was run by Bernard Lawrence Madoff, worth about $64.8 billion [13]. Bernard L. Madoff Investment Securities was started in 1960, but he testified that the scheme started in 1991. The scheme ran for around 17 years and defrauded thousands of investors for billions of dollars [14]. Madoff claimed to generate extended and regular returns through an investing strategy called a split-strike conversion. However, he deposited the investments into a single bank account to pay existing investors who wanted to withdraw. He used his friendship with wealthy, influential business people and got positive recommendations from them to attract more investors [15]. Madoff tried to make the scheme prestigious by not accepting everyone to invest. The global financial crisis of 2008 made the scheme collapse [16]. In 2009, Madoff confessed to his sons, who turned him in. He pleaded guilty and was sentenced to prison and ordered to forfeit $170 billion in assets. He used to attract investors by a reputation for unbelievably high returns and refining his victims by earning their trust.

2.2 High Yield Investment Program (HYIP)

A high-yield investment program (HYIP) is an online version of a Ponzi scheme. It is an investment scam that promises an unbelievably high return on investment by paying early investors with the money invested by new incoming investors [17]. The U.S. Security and Exchange Commission defines *HYIPs* as "unregistered investments generally run by unlicensed individuals or businesses, often a fraud. The red flag of an HYIP scam is the promise of enormous returns at little or no risk to the investors" [18].

The fraudsters generally set up a website promising a very high-interest rate like 1-2% per day, disclosing almost no information about the underlying investments. Fraudsters may use social media or online forums for advertising an HYIP. They also encourage investors to use social media to share information about an HYIP with others and provide them a referral bonus in return [18]. The rise of digital currencies has made it much easier for operators of such websites to accept payments from anyone worldwide. It can be challenging to track down the actual account holders using social media due to potential anonymity, making it harder for fraudsters to be held accountable [19]. With cryptocurrencies, the HYIPs take advantage of their pseudo-anonymous behavior and prefer to use it to hide from law enforcement agencies. There are many active HYIP websites at any given time [20].

HYIPs have been around since people started using the internet widely. One of the reasons scammers can make HYIPs run efficiently is because of their popularity, making everything readily available. Kit developers like Goldcoders sell complete HYIP website kits. These websites can be easily set up and start running in a few minutes. The investment to promote HYIPs on various discussion forums and social media is entirely free of cost. Cryptocurrencies and payment processors like Perfect Money, not controlled by the government, make it easy for scammers to accept money for their investment schemes. The ease and accessibility of these resources have made it very easy for anyone with a meager investment to start an HYIP and continue the fraud by luring more and more investors.

2.2.1 Pirate@40 HYIP

Around September 2011, Trendon T. Shavers started an online bitcoin Ponzi scheme in the United States. He ran it as a classic Ponzi scheme over the internet, where the new investments are used to pay back the older investors. He was the founder and operator of Bitcoin Savings and Trust (BCS&T) that offered and sold Bitcoin-based investments through the internet. He promised his investors up to a 7 percent weekly rate of return [21]. Shavers was able to obtain approximately 146,000 Bitcoin in BCS&T investments as per the reports from the U.S. Department of Justice [22], but the precise number is unknown. The bitcointalk[1] forum estimates between 150,000 and 200,000 bitcoins. He claimed to support a Bitcoin market-arbitrage strategy. The scheme stopped running in August 2012 [23]. The President's financial fraud enforcement task force brought the charges while investigating financial crimes. Shavers pled guilty on September 21, 2015, and was sentenced to 18 months in prison [22].

[1] https://bitcointalk.org/index.php?topic=576337#post_toc_38

Chapter 3

Related Work

This chapter enlightens the readers with the significant contributions and background for identifying High Yield Investment Programs (HYIPs) through machine learning algorithms and aggregator websites. Related work on human factors identified influencing people to invest in Ponzi schemes and how these factors have evolved will be discussed.

3.1 Detecting HYIPs using Machine Learning

High Yield Investment Programs (HYIPs) running on cryptocurrency blockchains are not straightforward to identify by researchers due to the pseudo-anonymous and decentralized nature of blockchains. The Bitcoin and Ethereum pseudo-anonymity privacy model increases the challenge of investigating them. The operators and investors of such scams generally advertise these investment schemes in various cryptocurrency forums to attract users. One such known Bitcoin forum is *bitcointalk.org*[1].

In 2015, Vasek et al. scraped *bitcointalk.org* forum to study the various Bitcoin-based scams and identified HYIPs as one of the major frauds in the Bitcoin ecosystem contributing to more than $7 million [24]. Other researchers studying HYIPs followed the same approach and primarily collected data from *bitcointalk.org* forum which highlights how crucial this forum is for this field [25] [26] [27] [28] [29] [30]. Additionally in 2018, Bartoletti et al. have used *Reddit*[2]

[1] https://bitcointalk.org/index.php?board=207.0
[2] https://www.reddit.com/

to increase their dataset for identifying more HYIPs [25]. To investigate named addresses and transactions, researchers use *Blockchain.info*[3] which provides all transactions associated with a bitcoin address. However, Bartoletti et al. and Toyoda et al. do not analyze the dataset to study the lifetime, nor the total profit of Bitcoin HYIP scams that they identified [25] [27] [28] [29].

3.1.1 Detecting HYIPs in Bitcoin blockchain

With the growing amount of scams in the Bitcoin blockchain, many machine learning approaches have been experimented with to identify HYIPs. The Random Forest classifier has been found to produce better results, with the most significant feature being the frequency of transactions by Bartoletti et al. and Toyoda et al. [25] [27] [29]. One of the problems using supervised machine learning for this problem is class imbalance. The labeled training dataset has comparatively fewer HYIPs than non-HYIPs. This does not provide a good dataset for machine learning models.

Toyoda et al. in 2017 shift towards transactional pattern analysis rather than the conventional graph mining approach for identifying bitcoin addresses related to HYIPs [27]. The authors chose two strategies for pre-processing the data: address clustering (AC) to find addresses owned by the same entity before transactions retrieval and extract features from transactions without address clustering. It is worth noting that AC degrades the classifier's performance slightly and increases false positives. It is possible that address clustering reduces the sufficient population in the clusters degrading the performance. Thus, we should not use AC for HYIP identification. An efficient feature used is the frequency of signed integers assigned to each transaction based on the Bitcoin amount and pay-in or payout. This might be because the pay-in patterns for HYIPs are concentrated between a particular range and easily distinguishable from non-HYIPs. The model was successfully able to classify 83% addresses correctly with a false positive rate of around 4.4%. It might have been interesting to see the results of using AC with unsupervised learning.

In 2018, Bartoletti et al. acknowledged that there is a lack of Bitcoin HYIP

[3] https://www.blockchain.com/

datasets and that just the data collection from online forums like bitcointalk.org is not enough [25]. Some HYIPs provide a new address for every investor and so are not available on the online forums. The authors collect the data of some bitcoin addresses by registering in a few HYIPs, providing false data. However, this activity had not gone through the Institutional Review Board, which puts it in a gray area for being ethical or not, depending on when this research was conducted. Future research should focus on ethically collecting the cryptocurrency addresses to analyze such schemes.

Ignoring the research and outcomes from Toyoda et al. related to address clustering [27], Bartoletti et al. in 2018 use address clustering for their machine learning model. Toyoda et al. found that using address clustering degrades the classifier performance and increases false positive. However, the model designed by Bartoletti et al. still achieves a recall of 0.969 and can classify 31 out of 32 HYIPs. To improve the class imbalance problem, the authors have used a sampling-based approach modifying the distribution of instances. Another approach used is the cost-sensitive approach encoding penalties through a cost matrix, penalizing more for the misclassification of the minority class, i.e., HYIP. The results show that the cost-based approach provides a better result. However, as there are very few HYIPs, it would be good to see how the model performs with more HYIPs in the dataset, where there is a higher chance of penalizing.

In 2019, Toyoda et al. proposed a novel rate conversion method to avoid price volatility and the sampling technique reducing computation amount [29]. The authors have also scraped the internet to provide a larger dataset of around 2000 HYIP operator's Bitcoin addresses. Before feature extraction, Toyoda et al. converted the bitcoin transactions to a stable currency, which seemed to help avoid the bitcoin price volatility and is a straightforward approach. The model has been tested against the dataset provided by Bartoletti et al. in [25] which evaluates to detect $\sim$ 94% HYIPs successfully. It is impressive to see how a simple currency conversion could improve the model's accuracy. However, there is no mention of recall from Toyoda et al. As the model's testing dataset is minimal, it needs to be tested against

a more extensive dataset to find its exact precision and recall.

Toyoda et al. apply the concepts of anomaly detection using a time series analysis to find when an HYIP changes its characteristics [28]. The numerical features have been extracted from the transactions to obtain the time series vector for analysis. The authors introduce the Principal Component Analysis-based approach to calculate the anomaly score. There is no reason provided why the authors chose this approach and not any other anomaly detection technique. No other technique is tried or tested to compare the effectiveness. The authors tested their model against Pirate@40 HYIP and showed that they successfully detected the changing points like name updates and changes in the investment rule. However, it should be noted that the HYIP already stopped running long back, and so this cannot be called dynamic transaction analysis. The machine learning model's effectiveness should be verified by detecting anomalies on currently running HYIPs and predict when a running HYIP will change its features. Additionally, a possible future work could be to see if this anomaly detection technique can detect HYIPs in the Bitcoin blockchain ecosystem as HYIPs are not generally expected in a blockchain.

Toyoda et al., even after realizing that address clustering increases the false positives and decreases the model's performance [27], used it twice again in future research work [28] [29] and never considered running it with unsupervised learning.

3.1.2 Detecting HYIPs in Ethereum blockchain

Ethereum brings in new sets of challenges. Bartoletti et al. and Chen et al. collect the Solidity code of published smart contracts from *Etherscan*[4] to investigate the source code and identify HYIPs running on the Ethereum blockchain [31] [32] [33]. One of the biggest challenges Ethereuem brings is that most of the source code for the smart contracts running is not available. As per the data collection done by Chen et al., less than four thousand contracts have their source code available out of more than a million running on Ethereum blockchain [33]. Another limitation comes from etherscan API, which does not provide more than the last 10k transactions restricting data collection for contracts with more than 10k transactions. For

[4]https://etherscan.io/

a long-lived HYIP on the Ethereum blockchain, it is possible to have these many transactions.

Bartoletti et al., for the first time in 2018/19, investigated to detect HYIPs in the Ethereum blockchain [31]. They found only 184 HYIP Ethereum contracts, which are much less than Bitcoin or other running HYIPs. All available smart contracts were analyzed manually, showing that it is not that simple to detect HYIPs on the Ethereum blockchain. The authors devised a model with four requirements to identify an HYIP: (i) contract should be sending payouts to investors, (ii) the only pay-ins are by the investors, (iii) every investor makes a profit considering new investors continue to invest and (iv) risk for investors increases depending on the time they pay-in. The authors used the normal Levenshtein distance to identify the contracts whose source codes are not available with a value less than 0.35. There is no proper argument provided to support this 0.35 threshold. Although it is a helpful approach, one cannot say with 100% confidence if those are HYIP schemes. We can also interpret from the results that Ethereum only Ponzi schemes are not that popular.

To solve the issue in detecting Ethereum based HYIPs without the source code availability, Chen et al. use opcodes to detect these schemes [32] [33]. Furthermore, the authors have devised a machine learning model using XGBoost to detect HYIP smart contracts at the moment of its creation, the most significant feature being the opcodes [33]. However, the model proposed by Chen et al. lacks the experimental evaluation for this feature and hence cannot be presumed successful. The authors only present the results for detecting Ethereum HYIPs on the dataset compiled by Bartoletti et al. [31]. Their model detects 83% of those HYIPs even when considering a smart contract as HYIP with a predicted probability larger than 50%. The authors have disregarded the other nine Ethereum HYIPs, reasoning why they are not HYIPs trying to prove their model is 100% efficient. We can interpret that the machine learning model is ineffective with this threshold as it may have many false positives. The authors never mentioned any false positives and the recall for their model. Contradictory to Bartoletti et al. [31], Chen et al. estimate more than

400 HYIPs are running on the Ethereum blockchain [33], disapproving the earlier interpretation that Ethereum HYIPs are not popular.

Chen et al. improved their previous machine learning model and added new account features: mean, standard deviation, and skewness of difference of income and expenditure of investors [32]. Similar to the anomaly detection technique mentioned by Toyoda et al. [28], Chen et al. have also devised two approaches to identify HYIPs as anomalies by using the One-class SVM and Isolation Forest [32]. However, the experimental result presents that simply using Random Forest raises the precision to 95% and is better than XGBoost used by Chen et al. [33]. The most significant features are the three opcodes: SLOAD, AND, and CODECOPY. Furthermore, the analysis pointed out that HYIP detection cannot be seen as anomalies as the approaches have a very low recall. This means that using anomaly detection in the Ethereum ecosystem is not helpful. Another notable result is that the HYIP smart contracts account for 0.03% of all Ethereum contracts. This result strengthens the earlier interpretation we made from Bartoletti et al. results [31] that Ethereum HYIPs are still not very popular.

3.2 Using aggregator websites to collect HYIPs

Moore et al. in 2012 were one of the first few researchers to look into the aggregators and how they can be used to find HYIPs' lifetime [20]. Aggregators invest a small amount in an HYIP and keep tracking the payouts, which helps them present whether a scheme collapsed. Neisius and Clayton concluded that aggregators list HYIPs to earn money [34]. Therefore, the aggregators may keep displaying the HYIPs on their website even when they are not paying out to the investors to continue earning money through advertisements. However, this would give a bad reputation to the aggregator.

A bit contradictory to Moore et al. about the aggregators not colluding with HYIPs [20], Neisius et al. state that aggregators are paid to list HYIPs that are active [34]. This might be seen as providing a different aspect to the collusion results from Moore et al. A new insight provided by Neisius et al. is the developers provid-

ing software called 'kits' [34]. The kit is a low-cost ready-to-deploy software that anyone can easily set up to start an HYIP or aggregator business in a few minutes. To stop frauds like HYIPs, stopping kits' business would have a high impact because it will stop the ease of deploying an aggregator and an HYIP, making it more complex for operators to start and attract investors into such schemes. Although, another thought that comes is that if a business is closed, the other company will start dominating, as we saw in the case of the dark web marketplace, Silk Road. Targeting aggregators can help to a certain extent because there are many ways to advertise a business, like social media. The discussion comments analysis done by Moore et al. presents that the most significant users interacting are from the United States of America.

To find links between different frauds by Drew and Moore [35], they use the HYIPs collected by Moore et al. from *hyip.com* aggregator [20]. Drew et al. try to find similarities in the HYIP web pages by extracting the HTML tags and directory structure by a combined clustering approach to link different scam websites. We argue that this is not the best approach to link the HYIP scam websites because these websites are generally designed by the same kit developer [34]. However, the research by Drew and Moore was done almost during the same time when Neisius et al. worked to find the insights about the kit developers [34], and till then, nothing about kits were known. Our research verified that 233 out of 366 HYIPs are licensed by one kit developer - Goldcoders. Buying the HYIP website from the same kit developer would have similar HTML tags and directory structure due to the same developers developing the websites.

3.3 Factors affecting Human Behavior

Analyzing human behavior is essential to understand the investors' mentality and the Ponzi schemes and HYIP operators. It helps to explain the factors that the investors consider, if they do, before investing in a scheme. This section reflects on the research conducted in Ponzi schemes and HYIPs related to various human behavioral aspects.

Trust in referees Wilkins et al. in 2012 provide one of the first studies on the Ponzi scheme investor's decision-making process [36]. The authors surveyed 17 victims who invested in offline Ponzi schemes and presented one of the expected points about how the victims are angry at the fraudster and the recovery process after the operators are sentenced. The surveys' conducted by Wilkins et al. reveals that Ponzi schemes are spread mainly by word of mouth, and victims do not blame the people who introduced them to the scheme, i.e., the referees. This perspective is difficult to understand because even though the scammers disappear, the referees are the ones who introduced the victims to the scams in the first place.

The authors devised a function correlating various factors to a user being a victim to a Ponzi scheme [36]. The most significant one is *affinity and trust*. The research done by Amoah [37] seconds this. Others include investor's investment knowledge, risk appetite, investment company's failure awareness, understanding of Ponzi scheme, products they are investing in, and demographic factors like gender, educational and income level, marital status, and age group. The male-to-female ratio is more significant than one showing that males risk more than females by investing in such schemes. It also presents that most people were single and younger than 35. A straightforward interpretation can be that single people take more investment risks as compared to married ones. The income level shows that people with less income are more likely to invest in such schemes, showing that people want to get rich quickly. The educational level does not matter much as all kinds of educational level people became victims, showing education countermeasure is not the best way to solve such issues.

Lack of investment knowledge The surveys conducted by Wilkins et al. also present that people often disregarded the advice given to them by professionals and do not do the math to understand how the return rate per month adds up to the rate of return per year and is *too good to be true* [36]. We interpret this as people do not study the whole investment process and do not care about the underlying information about how it is possible to get such a large amount of return rate. The victims consider the received promissory note on a piece of paper as an official document,

which again confirms the amount of trust people have in the companies they invest in. The surveys also reveal how the victimization impacts the users personally due to the loss: guilt, shame, regret, financial loss, and depression, which is not only financially but also mentally frustrating.

Relying on aggregators In 2012, Moore et al. introduced a new human behavioral perspective that the investors understand the fraud and consciously decide to invest early in the chain to get a profit with a higher probability [20]. For the first time, research is conducted on online Ponzi schemes, often referred to as High Yield Investment Programs (HYIPs). The authors points out another stakeholder of the HYIP ecosystem called the aggregators. Investors can use aggregators to find HYIPs details and track whether it is paying out or not, raising the question of whether these aggregators collude with the HYIPs. The analysis presents that aggregators can collude with the HYIPs but generally do not, providing unbiased information to the public. However, only nine aggregators' data were collected in the research, and therefore, some aggregators not researched upon might be colluding with the HYIPs.

Future research should be done using only the aggregators who do not collude with HYIPs. However, it is not easy to find how many and which aggregators collude. There is also a possibility that the HYIP payouts to aggregators are more consistent than the investors to maintain the HYIP ratings on aggregators' websites. This is a kind of unconscious collusion as aggregators do not willingly collude. Another insight is that around 600,000 people search for HYIPs each month. There is no source referenced for these results, but in our research, we show a significant number of HYIPs still running in today's era.

Education/Qualification level Amoah provides an empirical understanding of factors that put investors at risk [37]. The author designed a logistic regression model to assess the chances of investors falling prey to such schemes. One of the factors - *education level* - according to the research conducted by Amoah, is supposed to significantly affect the chances of becoming a victim of Ponzi schemes. In contrast to this, the surveys conducted by Wilkins et al. in 2012 showed that even educated

people fell into such schemes [36]. Amoah also introduces some psychological aspects like *cognition gullibility* and *confirmation bias* explaining why people invest in such schemes, which should be studied further and compare those factors with today's investors in HYIPs to find the similarities and differences.

Observing a scheme before jumping in Vasek and Moore analyzed that the median time is about five days for victims to comment on a post after the initial advertisement of a scheme [30]. This result shows that investors take their time before they jump into new schemes, and users do not notice them immediately. The results by Vasek et al. present that a few new victims commented half a year after the initial advertisement. It points out that users follow other investors' activity before investing or simply that they were just new users. This is similar to what Wilkins et al. mentioned for offline Ponzi schemes in [36], where people wait for some time before investing in a scheme and observe if others are investing.

Age Boshmaf et al. in 2020 are the first ones to deeply analyze a single HYIP called MMM [26]. No previous work in the field has done an in-depth analysis for a single HYIP. MMM is one of the biggest HYIPs in history that operated for over five years. Analysis of the user profiles posted on the MMM's bitcointalk.org post reveals that the user's average age is 32.85 years old, which is very similar to the outcomes of the surveys conducted by Wilkins et al. [36]. It can be interpreted that most users would probably be between 25 - 40 years, showing that most older people are not a part of such schemes. This shows that people in their middle ages are the ones who are attracted to these get rich quick schemes and become the victims of such frauds. Furthermore, the authors also found that the users' profiles' locations reveal Indonesia, China, and India as the most popular countries where people interacted with the MMM post. This is a new insight compared to the finding from Neisius et al., where the majority of interactions were from the USA [34]. The gender ratio presents similar conclusions as to the surveys by Amoah [37] and Wilkins et al. [36] for Ponzi schemes.

HYIP website appearance and features Chiluwa et al. studied the content of the Nigerian HYIPs' web pages. They reflected how the content is explicitly created

such that it targets a particular audience [38]. The significance of the color attracts users from a specific country or region. The text generally shows a fake motive behind a scheme, and the designs appeal to the underprivileged. Operators have become more intelligent over the years and study the country and people they are targetting. This can even be interpreted as targetted HYIPs where websites are designed to attract a investors from a particular community. Fig 3.1 and Fig 3.2 show how operators present security features and try to lure investors by gaining user's trust.

Future work should also aim to understand the social science aspect of the HYIP schemes to understand the various human elements unknown to the research landscape. Xia et al. present insights into how scammers and fraudsters take advantage of naive people and use any opportunity they can to lure people into investing in such schemes [39]. Xia et al. discovered nine HYIPs brought into the environment during the COVID-19 pandemic and how scammers under the shadow of medical research try to run such schemes. However, the authors identified HYIPs using keywords in various security reports and Google, which is not the best approach to find HYIP scams.

Figure 3.1: HYIP macbulls.com luring investors by displaying features like licensed script and registered company.

It has not been possible to survey the fraudsters directly yet. Many Ponzi schemes and HYIPs have been successfully shut down by the government regulating agencies. Future works should focus on getting permission from the government and interviewing the operators of such schemes to provide insightful information in this field.

Figure 3.2: HYIP genessis.net luring investors by displaying features like DDOS guard and green bar SSL.

3.4 Factors affecting HYIPs lifetime

Moore et al. analyzed the data provided by the aggregators and used it to calculate HYIPs' lifetime [20] showing the survival analysis for the lifetime of the HYIPs. The survival function devised by the authors points that the median lifetime of HYIPs is just 28 days presenting that the lifetime of HYIPs is not very long. One in four HYIPs last for greater than three months, and one in ten lasts for more than ten months. However, this is not very accurate because aggregators also take time before trusting an HYIP, and this lifetime calculation would not add those days or weeks. Moore et al. present that around 80% HYIPs last a few days. For the remaining 20%, there is a different result from different aggregators where some show a longer life than others. There are various interpretations for this. One could be that the HYIPs identify the aggregators and only payout the ones from where more investors came to their scheme. This can be seen as unconscious collusion as aggregators might have no idea about this; however, there is no data to support this.

Rate of return Moore et al. also pointed out that a less generous profit rate results in a greater chance for HYIPs' more prolonged survival [20]. This means that investors understand scams and do not risk when they see a very high rate of return. Additionally, the *minimum investment period* required by an HYIP is concluded as a factor affecting the HYIPs' lifetime. This might be because investors do not want to be tied up for a long time in a particular scheme. Moore et al. also present a more common scenario: the users' ratings cannot predict the collapse because investors

keep pushing and posting good reviews about HYIP, hoping to get their profits. In contrast, an *aggregator's reviews* would have a better chance to present accurate results because they invest a minimal amount to keep a check if the HYIP is paying or not. Aggregators are generally not in the environment for profits from their investment.

Comments and post interactions Vasek et al. introduced a few factors affecting the lifetime of HYIPs: (i) scammer interaction with investors on the forums, (ii) 'shills" comments on the scam advertisement post in the forum, and (iii) number of days between a scammer registers the account on the forum and advertises the scheme [30]. The authors pointed out that the more scammer postings, the longer the lifetime of the schemes. The increase in the lifetime might be because more postings attract more investors and show that the scheme is actively operating. An average scam where the operator posted at least half of the posts lasted for about three weeks. Out of 1780 scams, half only lasted a week, and about a quarter of them did not even last a day. The authors define 'shills' as the victims posting only about a single scheme and nowhere else on the forum. This shows the investors who invest in one scheme and continue posting about it until they do not get back their investments and some kind of profit.

Vasek et al. also present that more shill posts increase the lifetime comparatively more than only scammer's posts or a combination of shills and scammer's posts. Investors posts may have a greater impact than the scammer's post in assuring future investors. However, the lifetime is close but not accurate as the authors only consider the difference between the post published and the last comment on the post. There is a possibility that the scammers shift the scheme's post to another forum after some time. It also determines that the scams with only one scammer's post have a lower lifetime than scammers that post on more scams.

Difference in days between opening an account and advertising a scheme Another exciting feature discovered by Vasek et al. is that the scams posted on the same day as scammers register their accounts are shorter-lived [30]. This might mean that investors believe in experienced users' schemes, where experience might

mean the days since a user is registered in the forum. It brings an excellent perspective about the human interaction needed to run scams, and investors do not like spam. The other way to see this is that desperate scammers do not survive for long. The authors devised the Cox proportional hazard model to refine the varying effects on HYIP lifetime. The model reflected that schemes with more active participation are longer-lived. A shill post relates to a 57% decrease in hazard rate showing that the victim posts attract new investors. At the same time, a newly registered account advertising scam has a 45% increase because investors do not easily trust a new account posting in the forum.

Openness of Ethereum Bartoletti et al. present a concept about investors' believing in Ethereum based HYIPs because the source code of smart contracts is public, stable, and automatically executed [31]. This confers that luring investors into schemes can be easy due to the openness of new technologies like Ethereum. Contradictory to this, Bartoletti et al. show that 60% of the HYIPs on Ethereum have a lifetime close to zero days.

3.5 Summary

As discussed, various machine learning techniques have been tried to identify and detect HYIPs. Researchers have also find other methods of collecting data from aggregators' websites with quite interesting factors affecting the lifetime of HYIPs and the investors. Factors like rate of return and post interactions have been found to affect the lifetime of the HYIPs. Surveys and empirical analysis conducted by various researchers have provided factors like trust in referees, age, and education level affecting human behavior influencing investors to invest in Ponzi schemes and HYIPs.

The previous research has provided various helpful ways to collect data, like using aggregator websites that we use in our research. The survival analysis has provided us a way to understand the lifetime of the HYIPs in the current era. A research gap exists in finding how HYIPs are still running and registering as a limited company. It is possible that this exploitation has recently started and was not

prevalent earlier, which makes it more of a reason for us to look into this. There is also a gap in exploring other factors that scammers use to advertise their schemes, like the various social media platforms. Section 4 explains how we collect the HYIP data and identify more variables used for analysis in Section 5.

Chapter 4

Methodology

This section provides an overview of the data collected and explains the technical methods used in the data collection of HYIPs and registration details. We perform data cleaning, sanity check of the addresses collected, and demonstrate how we plot this on a map using Google Earth. Lastly, we provide insights into the various collected variables like HYIPs registered in the UK, on which we perform analysis in Section 5.

4.1 Data Collection

We aim to measure the High Yield Investment Programs (HYIPs) by collecting data from one of the most popular aggregators *hyip.com*[1] where the high yield investment schemes/scams are advertised. Aggregators are services that monitor every individual HYIP, lists them on their website, and promote them [34]. While manually visiting the aggregator website, we discover a review forum[2] on *hyip.com* that provides individual threads for every HYIP. Fig 4.1 provides a screenshot of how the HYIP review forum on *hyip.com* looks like. The individual threads of an HYIP, as shown in Fig 4.2 and 4.3, entails detailed information about the investment scheme, updates on the return on investment, and even allows investors to comment on the thread. Instead of redirecting directly to the HYIP website, it provides a link to the thread of that HYIP on another aggregator's website - *hyiprank.com*[3]. Fig 4.4 shows

[1] https://hyip.com/
[2] https://hyip.com/forum/review
[3] https://hyiprank.com/

the comments on the individual thread of an HYIP on *hyiprank.com*. It provides a referral link to the HYIP website with the referrer *HYIPDOTCOM*. From this, we can assume that *hyiprank.com* is managed by *hyip.com*.

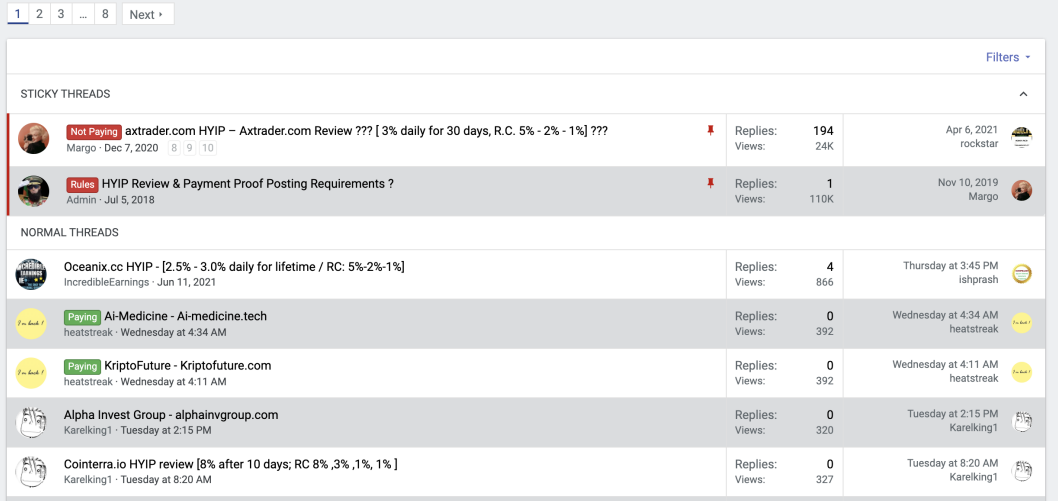

Figure 4.1: Screenshot of the HYIP Review Forum of the aggregator hyip.com.

Between November 11, 2020, and Aug 5, 2021, we made visits to the *hyip.com* review forum and the *hyiprank.com* aggregators almost every alternate day (with some exceptions and minor interruptions). We parsed the data after fetching from *hyip.com* review forum, to extract the URL for every individual HYIP thread on *hyip.com* and *hyiprank.com* and crawled them. We then parsed the individual threads on *hyiprank.com* and *hyip.com* to extract the HYIP website URL in order to crawl the HYIP web page. On April 16, 2021, the crawler was not able to run as usual. While manually investigating, we found that *hyip.com* changed their website design template. The review forum was shifted under another URL - *moneymakergroup.com*[4] and *hyip.com* still provides a link to the same through one of the tabs named as "Forum" on the top of the web page. Unsurprisingly, on April 19, 2021, just three days later, the *hyiprank.com* aggregator website stopped working, and the domain was not reachable. After 2-3 days, we checked again, and the domain started redirecting us to *hyip.com* and does so to date (last checked September 1, 2021). This supports our assumption made in the beginning that *hyip.com* and *hyiprank.com* were being managed by the same entity.

[4] `https://www.moneymakergroup.com/forums/review/`

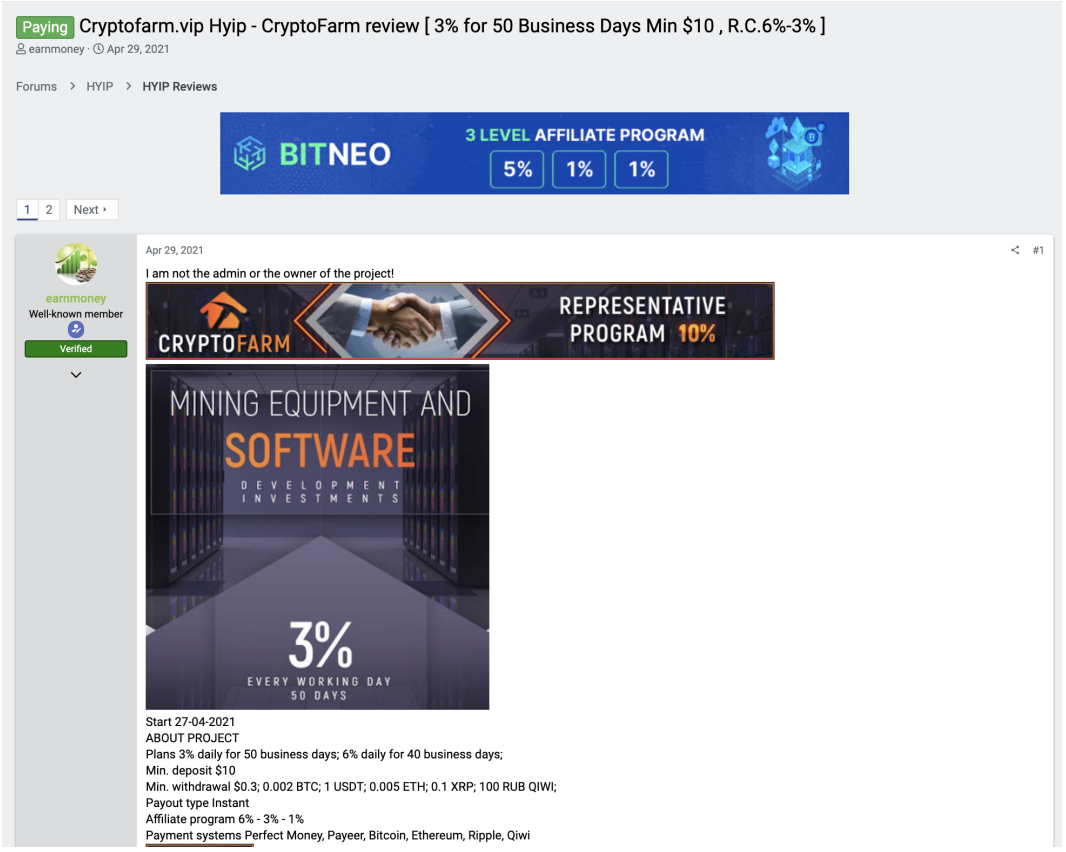

Figure 4.2: `https://cryptofarm.vip/` HYIP's individual thread on HYIP Review Forum at *hyip.com*.

4.1.1 Technicalities in Data Collection

All the automated visits to the aggregators' websites and the HYIP web pages were made using selenium[5] through different user agents, picked at random in the code as shown in Appendix A.1. We use selenium to fetch the website's source code once it is JavaScript has loaded completely. A 'user agent' is a character string that tells the server and network peers which browser and operating system makes a particular HTTP request. We made a list of possible agents as shown in 4.1 and pseudo-randomly chose one every time. This helps to ensure that the aggregators' websites and HYIPs would not identify the operating system or the browser being used to crawl the data. This practice helps us disguise and not let the server know that the same browser and operating system combination makes the same HTTP

[5] `https://www.selenium.dev/selenium/docs/api/javascript/module/selenium-webdriver/firefox.html`

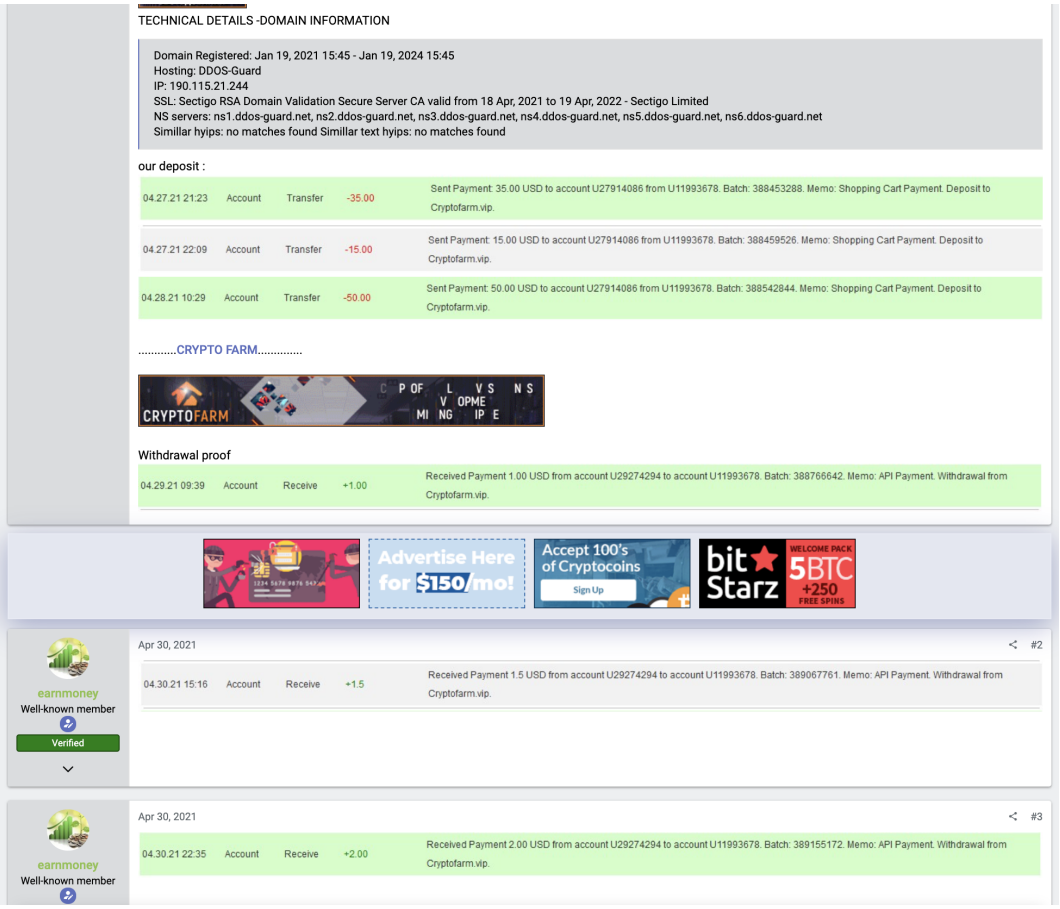

Figure 4.3: Comment section of `https://cryptofarm.vip/` HYIP's individual thread at *hyip.com*.

request every alternate day. Previous researchers used Tor[6] browser to hide their institution's identity or to avoid getting blocked by the aggregators, or HYIP websites [20]. We were able to collect the data without being blocked, and the data collection was not done on the university network. Therefore, we did not need Tor.

One of the challenges while crawling the HYIP websites is that they are listed on the aggregator's website and have an individual thread. However, the domain is either not reachable or redirects to some blogging or gambling website (a typical scenario where HYIP earns money even after their investment scam is over). In this case, we manually use *archive.org*[7] to find if any snapshots of the website are available so we can use that to crawl the web page, else, we do not consider that HYIP in our dataset. Some cases also include where *archive.org* instead of having

[6] `https://www.torproject.org/`
[7] `https://archive.org/`

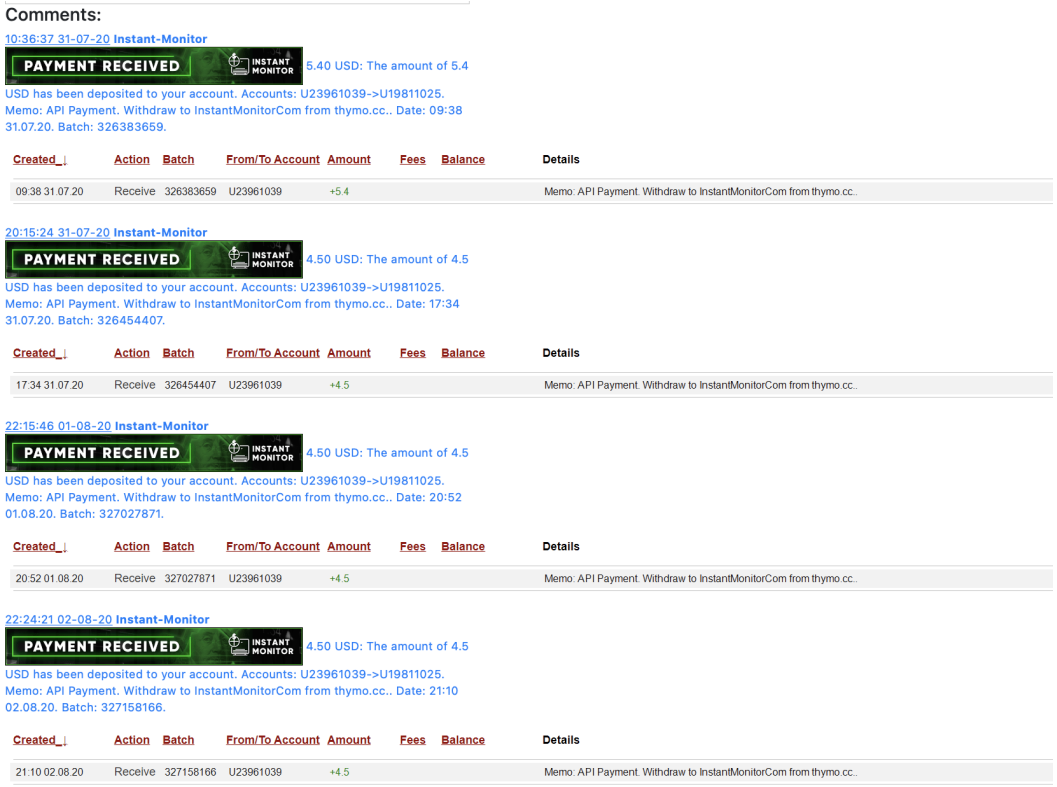

Figure 4.4: Comment section of `https://thymo.cc/` HYIP's individual thread at hyiprank.com.

snapshots of the web page, stores snapshots only of "Security check" pages like checking for a bot/actual human. In such cases ($\sim$ 17), we ignore the HYIPs due to a lack of evidence of existence.

'Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:82.0) Gecko/20100101 Firefox/82.0'
'Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:82.0) Gecko/20100101 Firefox/82.0'
'Mozilla/5.0 (X11; Linux i686; rv:82.0) Gecko/20100101 Firefox/82.0'
'Mozilla/5.0 (Linux x86_64; rv:82.0) Gecko/20100101 Firefox/82.0'
'Mozilla/5.0 (X11; Ubuntu; Linux i686; rv:82.0) Gecko/20100101 Firefox/82.0'
'Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:82.0) Gecko/20100101 Firefox/82.0'
'Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:82.0) Gecko/20100101 Firefox/82.0'
'Mozilla/5.0 (iPad; CPU OS 10_15_7 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) FxiOS/29.0 Mobile/15E148 Safari/605.1.15'

Table 4.1: List of user agents used at random in the code while collecting data.

4.1.2 Registration Details of HYIPs

Attaining a list of HYIPs, the next goal is to learn more about them and find the details and addresses they provide on their website to their users. Manually investi-

gating the HYIP websites, we found that most HYIPs mentioned their address and even provided registration details. This was surprising as the HYIPs conducting investment scams lured investors by proving that they are registered as a limited company by a government body in their respective countries. Every country has its respective government entity, which provides a company registration certification to a business in order for them to conduct business in their country legally. We parsed the crawled HYIP web pages to find and collect their details like the address, contact number, and registration number/document. We find that a majority of the HYIPs (230 out of 366) provided an address in the United Kingdom and even included their registration documents of being a 'limited company' in the United Kingdom.

Companies House is the government body that provides registration to a business as a 'limited company' in the UK. In the United States of America, the Secretary of the State provides testimony and registers a business as a limited company. Similarly, other countries have their governing bodies to register and conduct their business as a 'limited company.'

We used the *Companies House* API to collect data from the Companies House database using the registration number provided by the HYIPs [40]. Fig. 4.5 shows an original official registration document provided by the Companies House.

While investigating other HYIPs manually, we also found HYIPs that have photo-shopped the publicly available official certificate and made it look like they are registered as a company in the country. Fig 4.6 is an example of an HYIP kit developer using a fake registration document of the UK Companies House, which we found on an HYIP web page. A kit developer is a business that provides ready-to-use website templates/designs to HYIPs and aggregators [34]. It is also possible that *visualhyip.com* provides similar photo-shopped unofficial and illegal certificates to their clients who buy the templates from them. From this incident, we can assume that these company registration certificates play a role in luring investors into the scheme by providing them a fake sense of genuineness.

Some HYIPs were registered in other countries like the United States of America (USA), the British Virgin Islands, Australia, Hong Kong, the Republic of Mar-

CERTIFICATE OF INCORPORATION
OF A
PRIVATE LIMITED COMPANY

Company Number **672850**

The Registrar of Companies for Scotland, hereby certifies that

COINIZIE LTD

is this day incorporated under the Companies Act 2006 as a private company, that the company is limited by shares, and the situation of its registered office is in Scotland.

Given at Companies House, Edinburgh, on **2nd September 2020**.

The above information was communicated by electronic means and authenticated by the Registrar of Companies under section 1115 of the Companies Act 2006

Figure 4.5: Official UK Companies House registration certificate for Coinizie Ltd.

shall Islands, and the Republic of Seychelles. They even provided the certificate of registration as shown in Fig 4.7, Fig 4.8, Fig 4.9, Fig 4.10 and Fig 4.11.

Legally, all finance-related activities in the UK are supposed to be enforced and regulated by the Financial Conduct Authority (FCA). Similarly, the Securities and Exchange Commission (SEC) regulates all finance activities in the USA. This means that any company doing finance-related activities like investments needs to be registered and verified by the respected government authority and is not allowed to do so unless verified. To lure investors into their scheme, some HYIPs try to be more innovative and provide false claims of being registered by the SEC in the USA, as shown in Fig 4.12. If an HYIP is registered as a limited company in the United Kingdom, it should be verified by the FCA and not U.S. SEC. Otherwise, the HYIP needs to be registered as a limited company in the U.S., which is not the case here.

4.1.3 Data Storage

Crawling web pages and HYIP threads almost every alternate day consecutively for ten months from *hyip.com* and *hyiprank.com* aggregators and HYIP websites result in a few hundred Megabyte of data. Secondly, collecting data for research makes the researchers responsible for safely storing the data, so it does not end up in the wrong hands. We planned to use the UCL Computer Science department virtual machines, but they do not allow students to install new packages needed for our codebase. Therefore, we created a private repository on GitHub and stored the complete raw data gathered and the analyzed data along with the codebase in the repository. The raw data is not accessible to anyone except the student researcher and the supervisor.

4.1.4 Plotting HYIPs on a map registered in the UK

Finally, due to the majority of the HYIPs providing a UK address or UK Companies House registration detail or both, we are interested in finding whether the addresses provided by the HYIPs exist or not. This can answer quite a few things about how the scammers choose the addresses, which areas in the UK are mostly populated with such scams, and if any correlations exist. For the sanity check, we

are using *getaddress.io*[8] to get the available postcodes and addresses in every postcode. We then match these with the ones provided by the HYIPs in the Companies House documents and their websites. We also collect the latitudes and longitudes of the addresses through *getaddress.io*, label them and then gather them in a Keyhole Markup Language (KML) file and plot the coordinates on *Google Earth*[9]. Fig 4.14 provides the plot of all HYIPs that provided an address in the UK. We can see that most postcodes are concentrated in and around London, which can be seen in Fig 4.13.

Eleven out of 230 different postcodes were not available through *getaddress.io* API, and so we searched those postcodes manually on google to find their latitudes and longitudes. As per *doogal.co.uk*[10] nine of out of these eleven postcodes are no longer in use. This can be a reason why *getaddress.io* API did not respond or provide a null output for these postcodes queries. It should be noted that the latitudes and longitudes collected and marked on the map are not precise to the exact address we gathered, but what *getaddress.io* provides us, where the latitude and longitude are the same for all addresses in a particular postcode. However, it still provides an interesting and useful visual representation. This can be used to study further the patterns in addresses chosen by scammers in the UK.

4.2 Identified Variables for Analysis

For the time period between November 2020 and August 2021, the aggregators *hyip.com* and *hyiprank.com* lists more than 500 HYIPs. After checking their existential validity, i.e., making sure the HYIPs web page or a snapshot of the web page on *archive.org* exists, we gathered an impressive list of 366 HYIPs with their individual threads present on the aggregators' websites. The *hyip.com* aggregator contained a few HYIPs that did not have an individual thread on *hyiprank.com* aggregator and vice versa, although we still collected such HYIPs and added them to our knowledge base.

[8] https://getaddress.io/
[9] https://www.google.co.uk/intl/en_uk/earth/
[10] https://www.doogal.co.uk/

The data collected and used for analysis has been provided by the scammers themselves, and therefore it should be critically assessed. We have used the collected data to derive several key measurements as explained in this section and summarized in Table 4.2 used for analysis in Section 5.

Variable Name	Obs.	Lower bound	Upper bound	Mean	Std. Dev.
Valid UK Address	230	0	1	0.75	0.43
Registered in UK	230	0	1	0.89	0.31
Same UK Address	230	0	1	0.15	0.36
Valid Goldcoders license	366	0	1	0.64	0.48
Currencies used	366	1	14	5.71	2.56
Social Media Platform	366	0	5	1.18	1.35
Contain Contact Number	366	0	1	0.20	0.40
Lifetime (days)	366	0	2864	158.52	307.86

Table 4.2: Summary of variables collected for HYIPs that we use in the analysis.

4.2.1 Countries in which HYIPs resides

Several HYIPs identified provided an address on their web page. Using these addresses, we clustered HYIPs according to countries. Table 4.3 shows the distribution of HYIPs among various countries based on the addresses provided by the HYIPs on their websites. Companies House in the United Kingdom is a government body where every company in the UK must be registered and made available to the public. Similar authorities exist in various countries like Hong Kong, the Republic of Marshall Islands, the Republic of Seychelles, and others. The third column in the table states the number of HYIPs officially registered as a limited company with the respective countries' government bodies and even provides a certificate of registration on their websites.

Discussion on HYIPs' registrations As we can see from Table 4.3, 56.01% of the HYIPs analyzed are found to be registered as a company in the United Kingdom. Therefore, we will mainly focus on the Companies House in the United Kingdom, the official government body for registering a limited company in the UK. Companies House maintains a database/register for all the companies that want to conduct any business in the United Kingdom. The procedure to start a limited company in the United Kingdom is much simpler than one can expect. Trusting the documents

Country	# of HYIPs claiming to be from the country	# of HYIPs registered
United Kingdom	230	205
Hong Kong	7	7
Republic of Seychelles	1	1
Saint Vincent and the Grenadines	1	0
Republic of the Marshall Islands	2	2
United States of America	6	3
Australia	5	2
Germany	3	0
France	1	1
Netherlands	1	0
Belize	3	2
New Zealand	2	1
British Virgin Islands	3	1
Singapore	2	0
Not Reported	99	–

Table 4.3: HYIPs distribution in various countries. $N = 366$.

provided by the individuals who want to start a company, Companies House, being a government body, lacks statutory power or capability to verify the data provided while registering a company [41]. The only checks conducted are to make sure the documents are complete and signed. This provides an opportunity for a scammer to take advantage of this loophole and register an HYIP as a limited company with the UK Companies House and attain an official registration document which they use to lure victims into investing in their investment scams.

The Financial Conduct Authority (FCA) is the equivalent of the United States Securities and Exchange Commission (SEC) in the United Kingdom. FCA is responsible for verifying and conducting all regulated activities in the UK, including all finance management companies. The issue arises as the HYIPs register their company with the UK Companies House as a 'fund management company' or 'financial investment' or even as a 'bank,' which is enough to make a naive individual in believing that the website is a trustworthy investment scheme and not an HYIP scam.

Another possible reason for such companies continuing to register and run is a lack of information sharing between the FCA and the Companies House. Addi-

tionally, the Companies House provide any nature of business - Standard Industrial Classification (SIC)[11] of economic activities to the company on demand like 'fund management activities' without the company getting verified and validated by the FCA which should not be provided so quickly.

4.2.2 UK Address validation

Every HYIP that claims to be in the United Kingdom provides an address claiming to be their office address. Some only provide the Companies House registration number. We made a complete collection of addresses of HYIPs in the UK from the HYIP websites and the details HYIPs provide to Companies House while registering themselves as a 'limited company.' We then used these addresses to check if the addresses provided by the companies are valid or not where we used the *getaddress.io* API. We used this API because it contains up-to-date UK addresses for all postcodes and is easy to use, providing all information we need, including coordinates. We say an HYIP has a valid address only and only if the complete address along with the postcode provided by the HYIP exists in the *getaddress.io* API. If the HYIP does not provide a number or the numbering is not in the list of addresses in that postcode as per *getaddress.io*, we treat it as an invalid address.

We found that 172 HYIPs provide a valid address out of 230 HYIPs. We check this because we want to see if having a valid UK address increases or decreases the HYIP's lifetime, affecting its success. One HYIP (*bitero.io*[12]) changed their address after some time from a valid address to an invalid address. 28 HYIPs have the same valid address as some other HYIP. 6 HYIPs have the same postcode as some other HYIP. Overall, we find 34 HYIPs that use the same postcode as some other HYIP, which is another variable we use for analysis. This shows that HYIP operators use the same addresses for different schemes. It is also possible that the same operator runs different schemes under the same address or various operators use the same known addresses.

[11] https://www.gov.uk/government/publications/standard-industrial-classification-of-economic-activities-sic

[12] https://web.archive.org/web/20210101010539/https://bitero.io/

4.2.3 HYIP Goldcoders license

Goldcoders[13] is one of the most commonly used HYIP kit developers. We checked all the domains of the HYIPs we collected for their license on *Goldcoders* website. We automated this check, but the CAPTCHA on the Goldcoders' checkdomain URL[14] did not work with the optical character recognition code developed, forcing us to collect this data manually. We found that 233 out of 366 HYIPs have licenses from this specific HYIP kit developer, more than 63.6% of the total HYIPs we investigated.

4.2.4 Currencies accepted by HYIPs

With the rise in digital currencies and wallets, investment frauds have found more than one way to accept payments from their investors. After investigating all the HYIPs' websites, we present the distribution of various forms of e-currencies and cryptocurrencies that the collected HYIPs accept in Table 4.4. We used the HYIP aggregator threads that we scraped and then ran a simple grep command (`grep -n "Accepts:<" *`) to provide the list of payment processors that a particular HYIP accepts. We searched for other keywords: 'Payment Systems:,' 'Accept Payment:' and 'Payment Processors' with the grep command.

We manually went through the crawled HYIP web pages for those who did not have this information in the aggregator website's HYIP thread. Fourteen currencies are the maximum an HYIP was found to be accepting, and the minimum is one currency. This shows that some HYIPs provide limited payment processors, in which case they generally choose the most common ones like Perfect Money or Bitcoin. On the other hand, some HYIPs provide multiple options assuming that it would attract more investors as the victim can invest in the currency he/she prefers.

4.2.5 Social Media Platforms used by HYIPs

The Internet has given birth to a lot of social media platforms. People use these platforms to connect with other people. With the high presence of people on social media, companies have also started aggressively marketing their products. Like

[13] `https://www.goldcoders.com/`
[14] `https://www.goldcoders.com/?page=checkdomain`

Currency	# of HYIPs accepting the currency
Perfect Money	326
Payeer	191
Bitcoin	310
Bitcoin Cash	170
Litecoin	272
Dash	146
Ethereum	285
Dogecoin	174
Tether	33
Ripple	47
Monero	16
Advcash	12
Bank Wire	23
ePayCore	7
Yandex	7
Tron	29
Stellar	6
qiwi	9

Table 4.4: Distribution of various currencies used by HYIPs. $N = 366$.

companies, HYIPs also use these various social media platforms to publicize themselves, seeking new investors and convincing their investors by virtually "connecting" with their customers.

As mentioned above that we crawled all HYIP websites and stored them securely, we used this to automatically find keywords - `"t.me"` for Telegram[15] group chat, `"twitter.com/"` for Twitter[16] handles, `"facebook.com/"` for Facebook[17] pages/groups, `"instagram.com/"` for Instagram[18] handles and `"youtube.com/"` or `"youtu.be/"` for Youtube[19] videos/channels. We found 213 HYIPs have at least one social media platform used for advertising mentioned on their website and 9 HYIPs have all five social media handles. Table 4.5 highlights the distribution of social media platforms used by HYIPs. Overall we col-

[15] `https://telegram.org/`
[16] `https://twitter.com/`
[17] `https://www.facebook.com/`
[18] `https://www.instagram.com/`
[19] `https://www.youtube.com/`

lected more than 450 social media handles as some HYIPs have more than one Telegram, Facebook or Youtube handles.

Social Media Platform	# of HYIPs using the platform
Telegram	187
Facebook	84
Twitter	67
Instagram	31
Youtube	62

Table 4.5: Distribution of various social media platforms used by HYIPs. $N = 366$.

4.2.6 Contact Numbers provided by HYIPs

We collect all contact numbers provided by the HYIPs displayed on their website and add them to our dataset for analysis and future research. We used the scraped HYIP website pages to automatically collect the numbers by running a simple grep command searching for numbers like `"+44"` or keywords like `"tel:."` In total, 73 HYIPs provide a contact number.

4.2.7 Lifetime

In order to find the lifetime of HYIPs in days, we used the scraped data from the aggregators *hyip.com* and *hyiprank.com*. We find the start date that the aggregators mention in the individual HYIP threads. To find the last date that the HYIP was active on, we use the date of the last thread on the aggregators' website or the last date for which *archive.org* has the snapshot of the HYIP website (whichever is later). As some HYIPs have been still up and running, we use August 9, 2021, as the date for finding the lifetime for these running HYIPs as we cannot guarantee their lifetime after that date. We used August 9, 2021, since it is the last observation date in our samples.

Censoring: While finding the lifetime of HYIPs using the start dates and end dates, 44 HYIPs were found to be still up and running as of August 9, 2021. For all the running HYIPs whose collapse date is unknown, we apply Type 1 Censoring, where the lifetime is right-censored. We apply right censoring on subjects who are

still alive and their birth time is known, but either get lost or can not be followed up anymore because they are alive and the study finishes.

4.3 Ethical Considerations

We submitted an ethics application to collect all the data for the research project. We were provided a full ethics approval from the University College London Research Ethics Committee that is valid till August 6, 2022.

4.3.1 Reporting to Financial Conduct Authority

While conducting this research, we came across HYIP websites which are still up and running. Considering the Ethics Committee's suggestions, we reported all the running HYIP websites that claim to be registered in the UK to the Financial Conduct Authority (FCA). As we discussed in 4.2.1, all regulated financial activities in the UK are required to be approved and verified by the FCA before conducting any regulated business in the UK. FCA does not update the results of the investigations conducted by them. Hence, we cannot provide any further results on the reported HYIPs.

CERTIFICATE OF INCORPORATION
OF A
PRIVATE LIMITED COMPANY

Company Number **123456**

The Registrar of Companies for England and Wales, hereby certifies that

VISUALHYIP.COM

is this day incorporated under the Companies Act 2006 as a private company, that the company is limited by shares, and the situation of its registered office is in England and Wales.

Given at Companies House, Cardiff, on **21th February 2015**.

The above information was communicated by electronic means and authenticated by the Registrar of Companies under section 1115 of the Companies Act 2006

Companies House

REGISTRAR OF COMPANIES FOR ENGLAND AND WALES

THE OFFICIAL SEAL OF THE REGISTRAR OF COMPANIES

Figure 4.6: Photo-shopped UK Companies House registration certificate by `https://visualhyip.com/`.

編號 2853334
No.

公 司 註 冊 處
COMPANIES REGISTRY

公 司 註 冊 證 明 書
CERTIFICATE OF INCORPORATION

* * *

本 人 謹 此 證 明
I hereby certify that

ZION FINANCE LIMITED

於 本 日 根 據 香 港 法 例 第 622 章《 公 司 條 例 》
is this day incorporated in Hong Kong under the Companies Ordinance

在 香 港 成 立 為 法 團 ， 此 公 司 是 一 間
(Chapter 622 of the Laws of Hong Kong), and that this company is

有 限 公 司 。
a limited company.

本 證 明 書 於 二 O 一 九 年 七 月 十 七 日 發 出 。
Issued on 12 July 2020 .

香港特別行政區公司註冊處處長鍾麗玲
Ms Ada L L CHUNG
Registrar of Companies
Hong Kong Special Administrative Region

註 Note :
公司名稱獲公司註冊處註冊，並不表示獲授予該公司名稱或其任何部分的商標權或任何其他知識產權。
Registration of a company name with the Companies Registry does not confer any trade mark rights or any other intellectual property rights in respect of the company name or any part thereof.

Figure 4.7: Company registration certificate for an HYIP in Hong Kong.

REPUBLIC OF THE MARSHALL ISLANDS
OFFICE OF THE REGISTRAR OF CORPORATIONS

CERTIFICATE OF INCORPORATION

I HEREBY CERTIFY that

FUTURE TRADE INT LTD.
Reg. No. 103073

is duly incorporated and has filed articles of incorporation under the provisions of the Marshall Islands Business Corporations Act on

November 11, 2019

WITNESS my hand and the official seal of the Registry on November 11, 2019

Peggy Yeung
Deputy Registrar

REGISTRAR OF CORPORATIONS
SEAL
REPUBLIC OF THE MARSHALL ISLANDS

Figure 4.8: Company registration certificate for an HYIP in the Republic of the Marshall Islands.

Republic of Seychelles
INTERNATIONAL BUSINESS COMPANIES ACT, 2016
(Act 15 of 2016)

Certificate of Incorporation

THIS IS TO CERTIFY that, having satisfied all the requirements in respect of incorporation under the International Business Companies Act, 2016,

THE FUTURE TRADE INT LTD.

is incorporated in the Republic of Seychelles as an International Business Company,

on this 11th day of **December 2019**

Given at Victoria, Seychelles.

Company No: 217244

REGISTRAR OF INTERNATIONAL BUSINESS COMPANIES

Figure 4.9: Company registration certificate for an HYIP in the Republic of Seychelles.

Certificate of Registration of a Company

ASIC
Australian Securities & Investments Commission

This is to certify that

MINEXO PTY LTD

Australian Company Number 648 952 425

is a registered company under the Corporations Act 2001 and is taken to be registered in New South Wales.

The company is **limited by shares.**

The company is a **proprietary** company.

The day of commencement of registration is **the twenty-fourth day of March 2021.**

Issued by the
Australian Securities and Investments Commission
on this twenty-fourth day of March, 2021

James Shipton
Chair

CERTIFICATE

Figure 4.10: Company registration certificate for an HYIP in Australia.

STATE OF ARKANSAS

SECRETARY OF STATE

John Thurston
ARKANSAS SECRETARY OF STATE

To All to Whom These Presents Shall Come, Greetings:

I, John Thurston, Arkansas Secretary of State of Arkansas, do hereby certify that the following and hereto attached instrument of writing is a true and perfect copy of

Articles of Organization

of

ARBIRATE LLC

filed in this office
October 15, 2020

In Testimony Whereof, I have hereunto set my hand and affixed my official Seal. Done at my office in the City of Little Rock, this 15th day of October 2020.

John Thurston
Secretary of State

Online Certificate Authorization Code: 3626525f888d6bb47f2
To verify the Authorization Code, visit sos.arkansas.gov

SEAL OF THE SECRETARY OF STATE ARKANSAS

Figure 4.11: Company registration certificate for an HYIP in Arkansas in the United States of America.

Figure 4.12: Company claiming to be registered by the US Securities and Exchange Commission.

Figure 4.13: Google Earth plot of HYIPs addresses in London, UK.

Figure 4.14: Google Earth plot of HYIPs addresses in the UK. $N = 230$.

Chapter 5

Analysis

This section mainly focuses on analyzing the collected data with statistical methods and presenting exciting insights. We visualize the data with plots to understand the lifetime of the schemes, how the different variables mentioned in section 4.2 are correlated, and which variables affect the lifetime of High Yield Investment Programs (HYIPs).

5.1 Understanding the lifetime of HYIPs

We calculate the lifetime of all 366 HYIPs by collecting their start date and end date as mentioned in 4.2.7. For the ones that are still running and do not have an end date, we use August 9, 2021, and mark the value for the censored variable as 1. We use this date as the collection ended on this date, and we did not observe HYIPs after that. A survival function is used to measure the survival probability, i.e., the conditional probability of an HYIP surviving beyond that time, given that an HYIP has survived just before that time [42]. It is the probability that an HYIP's lifetime is greater than t days. We use the survival function $S(t)$ and visualize the data in Figure 5.1.

In order to get the proper survival function, we use Type 1 censoring. When an experiment has a fixed number of items, and the experiment is stopped at a predetermined time, any item remaining at this point is right-censored or Type 1 censoring [43]. We have 44 HYIPs that are still running. It is impossible to see when these running schemes collapse, so we use survival analysis, where these 44

running HYIPs are 'right-censored.'

Kaplan-Meier estimator [44] is used to estimate the survival function $S(t)$ using the observed lifetime data. We use this to measure the fraction of HYIPs that collapse after a given date. Figure 5.1 plots the lifetime of HYIPs in days with the overall survival probability. The dotted lines in the plot are the 95 % confidence interval. The survival function shows that the median lifetime of HYIPs is 565 days. 89% of the HYIPs survive for more than a year, whereas only 38.6% of the HYIPs run for more than two years.

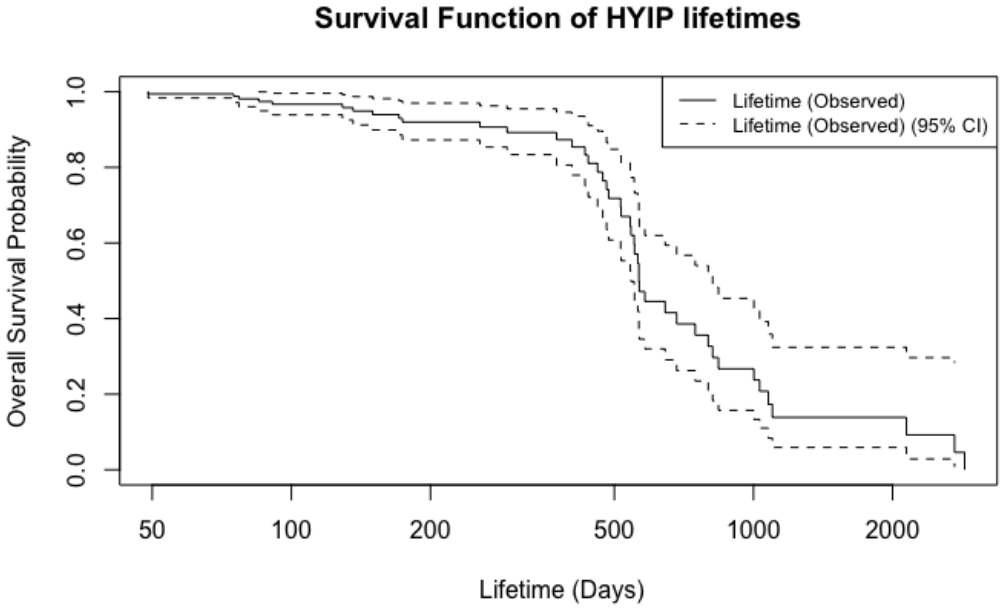

Figure 5.1: Survival Function of HYIP lifetimes. HYIPs $N = 366$.

The survival function for HYIPs claiming to be in the UK (HYIPs $N = 230$) has a median lifetime of 553 days. 86.4% of UK HYIPs survive more than a year, a little less as compared to all HYIPs. However, the 2-year survival probability is 38.5% which is almost similar to the probability of all HYIPs collapsing after two years. We see a significant difference for surviving more than 1000 days where the UK HYIPs have a probability of 16.5%, whereas the overall HYIPs show a probability of 26.7%.

The lifetime analysis by Moore et al. showed that the median lifetime of HYIPs is 28 days [20]. They also found that one in four will last more than three months,

and one in ten for more than ten months. In comparison, our research finds that the median lifetime of the HYIPs is 565 days and 89% of the investigated HYIPs survive for more than a year. A possible reason for this difference might be that Moore et al. collected the data 10 years ago, in 2010-11, when HYIPs were comparatively new. The operators at that time were scared to run it for a more extended period. Over the decade, the operators have learned a lot and improvised their schemes to make them run longer.

5.2 Variables affecting lifetime of HYIPs

The Log-Rank test is a method to check the hypothesis by comparing the survival distributions of two samples [45]. It tests a null hypothesis that there is no difference between the survival probability of the lifetime of the HYIP at any given time between the two groups. We examine all the variables collected independently with the survival function and use a log-rank test to confirm the statistical differences. For example, the test helps to check if the survival probability of the HYIP for accepting Bitcoin and not accepting Bitcoin at a given time shows a difference at a significant level. The test is likely to detect a difference between two groups when the risk of an event is consistently more significant for one group than another [46].

We study all variables mentioned in Table 4.2. For the variables social media platforms and currencies used, we individually use each social media platform summarised in Table 4.5 and similarly every currency summarised in Table 4.4. This gives us a better understanding of using individual social media platforms and payment processors instead of focusing only on these two broader variables.

We observe the significant difference in Ethereum and Litecoin for the survival function of all HYIPs lifetime. We show these in Fig 5.2 and Fig 5.3. The $p-value$ for Ethereum is $p = 0.0089$ and for Litecoin is $p = 0.065$ which shows that these are significant at 95% level and 90% level respectively. These small $p-values$ means the effect of Ethereum and Litecoin is important and helps to reject the null hypothesis. The red line shows the survival curve for the HYIPs that accepts Ethereum as a payment processor in Fig 5.2. The blue lines show the survival curve

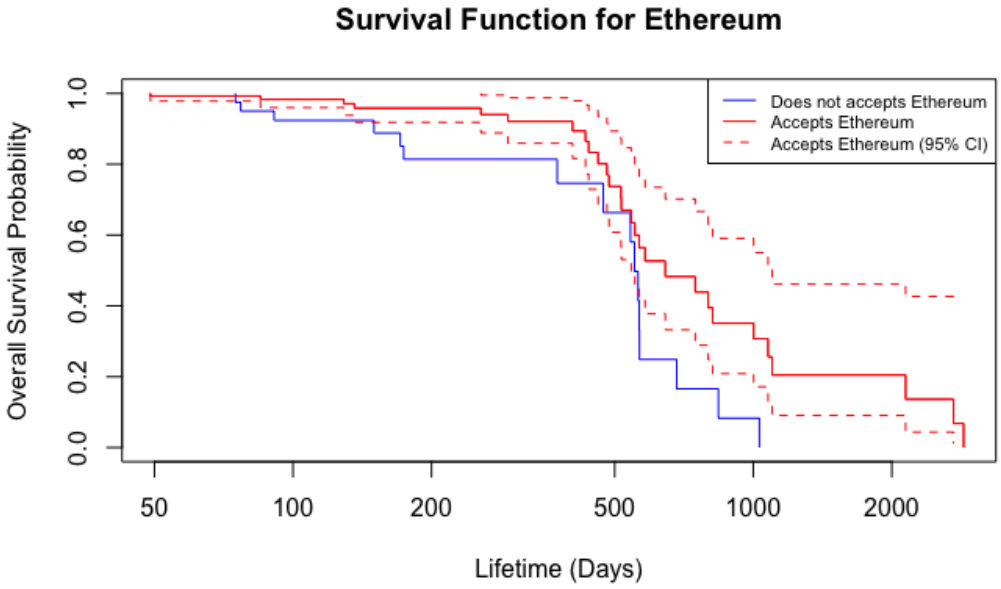

Figure 5.2: Survival curve of HYIP lifetimes for Ethereum. HYIPs $N = 366$.

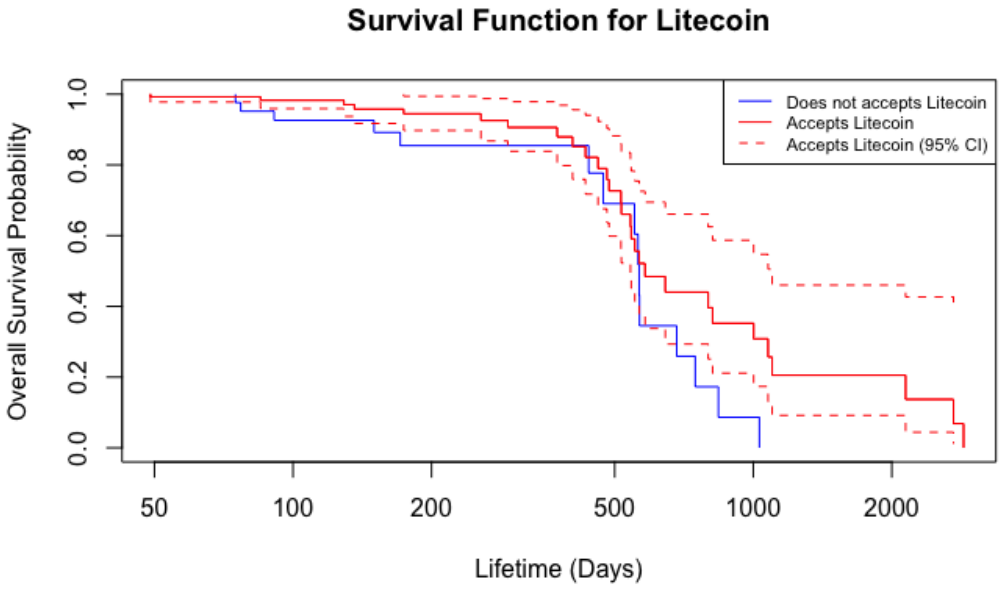

Figure 5.3: Survival curve of HYIP lifetimes for Litecoin. HYIPs $N = 366$.

for HYIPs that do not accept Ethereum. The HYIPs that accept Ethereum have a longer lifetime than the HYIPs that do not accept Ethereum.

Similarly, in Fig 5.3, the red line represents the survival curve for the HYIPs that accepts Litecoin as a payment processor, and the blue line represents the life-

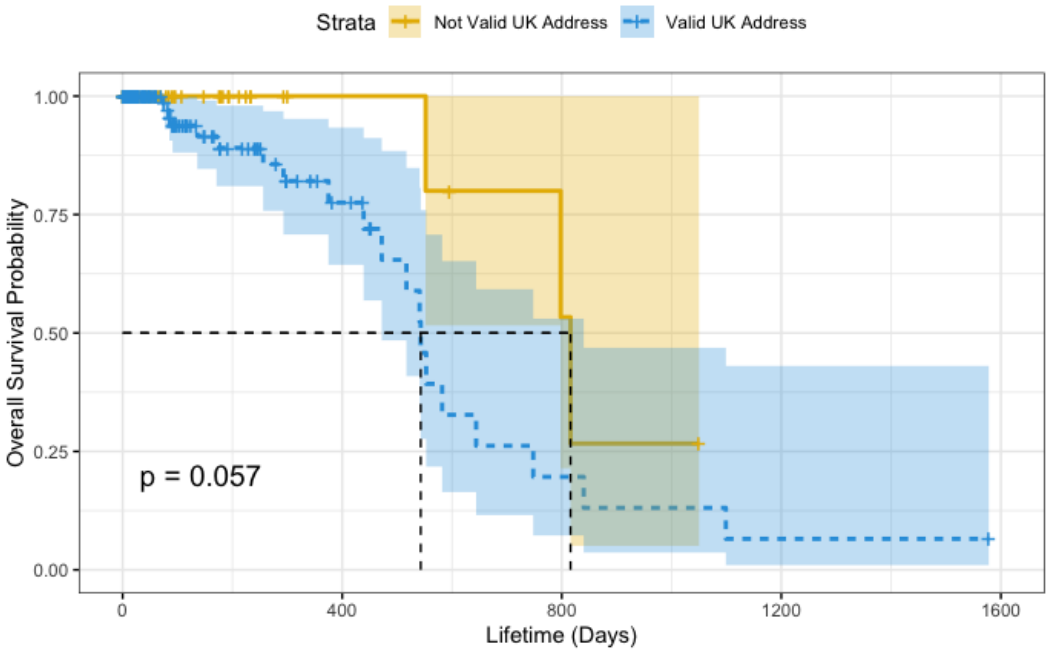

Figure 5.4: Survival curve of HYIP lifetimes for valid UK address. HYIPs $N = 230$.

time curve of the HYIPs that do not accept Litecoin. The HYIPs that accept Litecoin as a payment processor have a longer lifetime.

We find that the valid UK address variable shows a significant difference. For the valid UK address variable, the test gives a $p - value$ of $p = 0.057$ as shown in Fig 5.4 showing the effect is significant at 90% level. We can visually also notice that this variable shows the major difference for survival analysis.

The blue line in Fig 5.4 represents the survival curve for HYIPs that have a valid UK address, and the brownish line represents the lifetime curve for HYIPs that do not have a valid UK address. We find that HYIPs with a non-valid UK address have a longer lifetime. One of the reasons we can interpret for it is that most investors believe the given address on the HYIP websites and do not care enough to check the validity of these addresses before investing in the schemes. Another reason can be that our sample size is limited to 230 HYIPs claiming to be in the UK. It is also worth noting that some postcodes used by the HYIPs that we consider invalid are because they are not used officially anymore and were disregarded (discussed in Section 4.2.2).

In order to see how these variables are correlated, we use a correlation matrix

as shown in Table A.1. The correlation matrix has been computed using Pearson and Spearman correlations. We also visualize the correlation matrix of all variables in Figure 5.5. Positive correlations are displayed in blue color, and negative correlations are in red. The color intensity and the size of the circle are directly proportional to the correlation coefficients. However, it does not mean that all these correlation coefficients are significant.

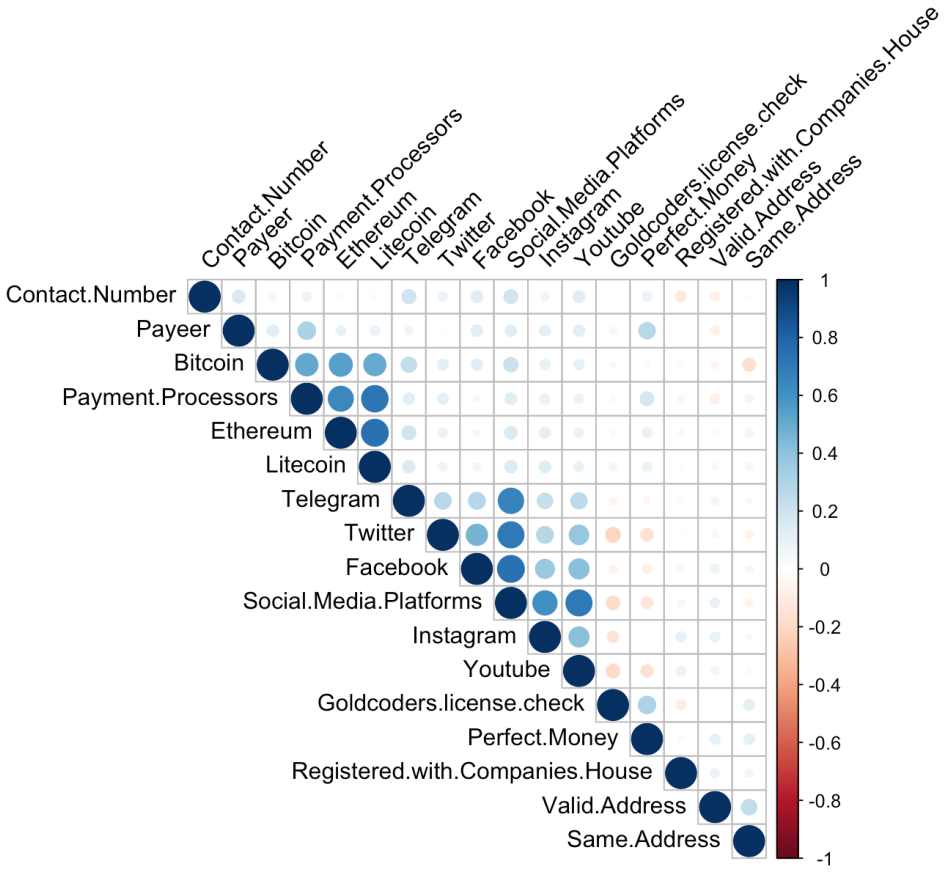

Figure 5.5: Correlation matrix plot of all variables. HYIPs $N = 366$.
The legend color shows the correlation coefficients and the corresponding colors.

5.3 Proportional Hazards Model

In order to observe the varying effects on the lifetime of the HYIPs, we run a Cox proportional hazards model. We use a hazard function $h(t)$, which describes the probability of an event or its hazard h if the subject survived up to that particular time point t [47]. We study how covariates are multiplicatively related to hazards,

both the dependent and independent variables. We will focus on the hazard rates for the Cox proportional hazards model, i.e., $exp(coef)$ column, and the $p-value$ for checking its significance. For the hazard rate, values greater than one correspond to an increase in the hazard rate, while those less than one correspond to a decrease. The hazard rate shows the probability that the HYIP will collapse, so an increased hazard rate means a greater risk of shutdown. In contrast, a decreased hazard rate means that the factor helps to increase its lifetime.

First, we run the Cox regression model on all UK HYIPs with all the variables including valid UK address, same UK address, and registered in UK as shown in Table 5.1. We see a 294% increase in the expected hazard rate using Twitter as a social media platform. This means that using Twitter as a social media platform for advertising the HYIP increases the risk of the HYIP collapsing by almost three times. One might expect that having a social media presence would increase the reach of the HYIP, attracting investors, and hence the HYIP should be longer-lived, but the opposite is true for using Twitter. Using Twitter as a social media platform decreases the lifetime of the HYIP. The effect of this is statistically significant at the 90% level, having a $p-value = 0.0836$. One of the reasons can be that the audience on Twitter is much more aware of the investment scams and might be reporting such accounts. It also shows that HYIPs might have a better chance to live longer by just targetting their usual investors who use aggregators to find HYIPs to invest in.

Previously in Table 5.1, we use all binary variables. Nevertheless, we also want to see the effect of HYIPs' numbers of social media platforms and the collective number of payment processors that HYIPs accepts. So we run another Cox regression on UK HYIPs as shown in Table 5.2. However, no variable has a statistically significant effect, as can be seen from this regression table.

Next, we run a Cox regression on all 366 HYIPs with all variables except the UK-only variables. We see from Table 5.3 that there is a massive 77% and 69% decrease in the expected hazard rate, respectively, by accepting Perfect Money and Ethereum as payment processors. This result is somewhat counterintuitive; using Perfect Money and Ethereum as a payment processor decreases the HYIPs risk of

Variable Name	$coef$	$exp(coef)$	95%CI	$p-value$
Contain Contact Number	0.72646	2.07	(0.59, 7.19)	0.2532
Payeer	0.79532	2.21	(0.56, 8.80)	0.2585
Bitcoin	1.29106	3.64	(0.38, 34.45)	0.2604
Ethereum	-0.78113	0.46	(0.08, 2.50)	0.3672
Litecoin	0.08575	1.09	(0.16, 7.43)	0.9303
Telegram	-0.73661	0.48	(0.16, 1.46)	0.1968
Twitter	1.37238	3.94	(0.83, 18.67)	0.0836 *
Facebook	-0.76182	0.47	(0.09, 2.47)	0.3705
Instagram	-0.22980	0.79	(0.08, 7.41)	0.8401
Youtube	-0.83125	0.43	(0.08, 2.48)	0.3485
Valid Goldcoders license	0.41737	1.52	(0.43, 5.41)	0.52
Registered in UK	-0.36832	0.69	(0.12, 3.95)	0.6787
Valid UK Address	0.71884	2.05	(0.62, 6.79)	0.2390
Same UK Address	0.63897	1.89	(0.35, 10.24)	0.4580
Concordance = 0.636 $(se = 0.069)$				
Likelihood ratio test = 12.76 *on* 14 df, $p = 0.5$				
Wald test = 11.14 *on* 14 df, $p = 0.7$				
Logrank test = 12.58 *on* 14 df, $p = 0.6$				
** Significant at the 95% level				
* Significant at the 90% level				

Table 5.1: Cox proportional hazards model: measuring all variables on the lifetime of UK HYIPs. $N = 230$.

collapse, prolonging the HYIPs lifetime. This means that using these two payment processors for accepting investments may play a significant role in increasing the lifetime of the HYIPs, getting more victims to invest in, hence increasing profits for the scammers. This prolonging lifetime effect of accepting Perfect Money is statistically significant at the 95% level on the lifetime of the HYIP, and that of accepting Ethereum is significant at the 90% level. One of the possible reasons for accepting Perfect Money having a comparatively better effect in increasing the lifetime of HYIPs might be because Perfect Money being an old payment processor, has a big enough community already using it, making it easy for investors to invest in the schemes.

Lastly, we also want to see the effect of HYIPs' number of social media platforms and the collective number of payment processors that HYIPs accepts. So, we run another Cox regression using these variables on all HYIPs. We can see from

Variable Name	$coef$	$exp(coef)$	$95\%CI$	$p-value$
Contain Contact Number	-0.10636	0.90	(0.36, 2.25)	0.820
# of Payment Processors	-0.02722	0.97	(0.78, 1.21)	0.808
# of Social Media Platforms	-0.13053	0.88	(0.65, 1.19)	0.397
Valid Goldcoders license	0.23070	1.26	(0.47, 3.39)	0.648
Registered in UK	-0.34917	0.71	(0.13, 3.72)	0.681
Valid UK Address	0.46517	1.59	(0.58, 4.38)	0.367
Same UK Address	-0.39691	0.67	(0.17, 2.70)	0.576
Concordance = 0.592 $(se = 0.068)$				
Likelihood ratio test = 2.59 *on* 7 df, $p = 0.9$				
Wald test = 2.52 *on* 7 df, $p = 0.9$				
Logrank test = 2.56 *on* 7 df, $p = 0.9$				

Table 5.2: Cox proportional hazards model: measuring the acceptance of number of social media platforms and payment processors along with other variables on the lifetime of UK HYIPs. $N = 230$.

Variable Name	$coef$	$exp(coef)$	$95\%CI$	$p-value$
Contain Contact Number	0.6767	1.97	(0.78, 4.95)	0.1507
Payeer	0.4639	1.59	(0.65, 3.90)	0.3106
Bitcoin	-0.1012	0.90	(0.28, 2.88)	0.8640
Ethereum	-1.1773	0.31	(0.08, 1.16)	0.0816 *
Litecoin	0.1713	1.19	(0.31, 4.53)	0.8020
Perfect Money	-1.4501	0.23	(0.06, 0.87)	0.0308 **
Telegram	0.1008	1.11	(0.48, 2.54)	0.8125
Twitter	0.8105	2.25	(0.76, 6.67)	0.1440
Facebook	-0.6223	0.54	(0.17, 1.69)	0.2881
Instagram	0.3538	1.42	(0.30, 6.74)	0.6555
Youtube	-0.5052	0.60	(0.18, 1.98)	0.4044
Valid Goldcoders license	0.661	1.94	(0.86, 4.36)	0.1092
Concordance = 0.61 $(se = 0.072)$				
Likelihood ratio test = 14.05 *on* 12 df, $p = 0.3$				
Wald test = 14.06 *on* 12 df, $p = 0.3$				
Logrank test = 15.03 *on* 12 df, $p = 0.2$				
** Significant at the 95% level				
* Significant at the 90% level				

Table 5.3: Cox proportional hazards model: measuring all variables on the lifetime of HYIPs. $N = 366$.

Table 5.4 that there is a 9% decrease in the expected hazard rate for the number of payment processors accepted by the HYIPs, significant at the 90% level. This means that the number of payment processors accepted by HYIPs may play a role in prolonging the lifetime of the HYIPs. Straightforward reasoning might be that

the more the number of payment processors a scheme accepts, the more ways an investor can invest into the scheme as per their ease, prolonging the lifetime of the HYIPs.

Variable Name	$coef$	$exp(coef)$	95%CI	$p-value$
Contain Contact Number	0.03963	1.04	(0.50, 2.17)	0.9156
# of Payment Processors	-0.09845	0.91	(0.81, 1.02)	0.0908 *
# of Social Media Platforms	0.01375	1.01	(0.81, 1.27)	0.9036
Valid Goldcoders license	0.26224	1.30	(0.68, 2.50)	0.4307
Concordance = 0.625 $(se = 0.054)$				
Likelihood ratio test = 4.33 on 4 df, $p = 0.4$				
Wald test = 4.28 on 4 df, $p = 0.4$				
Logrank test = 4.32 on 4 df, $p = 0.4$				
** Significant at the 95% level				
* Significant at the 90% level				

Table 5.4: Cox proportional hazards model: measuring the acceptance of number of social media platforms and payment processors along with other variables on the lifetime of HYIPs. $N = 366$.

The survival analysis and Cox proportional hazards model confirm with the available data that being registered as a limited company in the United Kingdom is not a factor affecting the lifetime of the HYIPs. A possible reason for this result might be the sample imbalance as 205 out of 230 HYIPs we identified in the UK are registered as a limited company. It is also possible that the investors do not care whether the HYIP is officially registered as a company or not. Collecting more data and looking into overall HYIP registrations outside the UK might provide new insights or strengthen the same result.

HYIPs having the same UK address also does not affect the lifetime. We would have expected such Ponzi schemes to be short-lived as they use the same address as another HYIP. Straightforward reasoning for this is that the investors do not care to check if the address is/was being used by someone else. However, this does not finish the research in this area. More factors need to be identified, like the various languages in which the HYIP website content is available and the default website language. This research would help identify the target victims/countries and which ones are more long-lived.

Chapter 6

Conclusion

High Yield Investment Programs (HYIPs) are one of the most significant existential financial frauds [1]. We collected a new data set of 366 HYIPs from November 2020 to August 2021. The data set presents new insights into registration details of HYIPs, addresses, contact numbers, social media handles, and various currencies used by these HYIPs. The report also presents a geographical plot of all HYIPs claiming to be in the UK. It verifies that most HYIPs that claim to be in the United Kingdom are registered as a limited company through the Companies House register. We analyze and show that valid UK addresses, accepting Litecoin, and Ethereum are three individual variables affecting the lifetime of HYIPs through survival analysis.

The Cox proportional hazards model helps to identify various variables having hazard rates with a statistically significant effect on the lifetime of the HYIPs. Social media platforms do not help in prolonging the HYIP lifetime at any significant level. Instead, we found that for UK HYIPs, using Twitter increases the risk of collapse almost by three times. We also see how using the number of payment processors increases the lifetime of HYIPs with a slight 9% decrease in hazard rate, simply because it provides multiple ways someone can invest in the schemes. We also see that Perfect Money increases the lifetime of the HYIPs with a statistically significant effect, primarily because of its ease of access and use. Accepting Ethereum decreases the hazard rate by 77%, and accepting Perfect Money decreases by 69%. Lastly, getting registered in the UK as a limited company does not affect the lifetime

of the HYIPs. This might mean that investors do not care if an HYIP is officially registered or not. It also means that HYIPs that register as a limited company are not better than HYIPs that do not register.

6.1 Recommendations

We check and confirm in this report that HYIPs exist, which register as a limited company in a few countries, including the United Kingdom and the United States of America. We provide the following recommendations:

Government organizations regulating finance-related activities We recommend that agencies like the UK Financial Conduct Authority (FCA), the U.S. Securities and Exchange Commission, and other similar organizations in their respective countries regulating finance-related activities collaborate with the organization that issues a company registration certificate. They should keep a check on businesses asking to be registered as a limited company and investigate if any of them want to conduct a regulated business activity.

Government organizations registering companies We recommend organizations like UK Companies House to verify the documents that the applicants submit before registering a business as an official limited company. We also recommend not to provide a Standard Industrial Classification of economic activities (SIC) code to a business without proper checks. Additionally, organizations like the UK Companies House should not be allowed to allocate codes like 'Banks' or 'Fund management activities' without the FCA verifying and approving the business conducting regulated activities.

6.2 Future Work

We made a collection of a lot of new variables and analyzed them. This does not still finish the work in Online Ponzi Schemes. Future research should focus on the following directions:

- *Collect more diverse data to strengthen our results:* As per our analysis, we saw that an HYIP being registered with Companies House in the UK does

not have any significant effect on HYIP's lifetime. As 89.13% of UK HYIPs are registered as a limited company, there might be no effect because of the unbalanced sample. We want to check the registration effect of all HYIPs in other countries and not limit it to the UK. It is possible that having more diverse data would show some change in results or confirm the result we got after the analysis.

- *Identify Factors affecting Human Behavior*: It is essential to find a more socio-technical analysis of why investors still invest in fraudulent schemes. There is a need for more updated surveys to study user behavior to understand the factors influencing their decisions to invest in HYIPs. Surveys should be conducted on HYIP victims and operators to understand human behavioral factors.

- *Collecting Operator's Digital Wallet Addresses*: Researchers should go through the ethical channel - Ethics Review Committee and collect various wallet/cryptocurrency addresses by registering in the identified HYIPs. Collected addresses should be analyzed, providing better insights into the money flow.

- *Exploring Ethereum Smart Contracts*: As discussed in Section 3, researchers have not been able to differentiate between the owners and users. Future research should discover ways to distinguish the investors' and scammers' earnings separately and present the victims' losses in numbers.

- *Identifying more variables*: We identified and collected a lot of new variables never looked into before by previous researchers for analysis. There still exist more variables that should be collected by investigating the HYIP websites and analyze how they affect the lifetime of the HYIPs. For, e.g., the default HYIP website language.

Bibliography

[1] Federal Bureau of Investigation (FBI). Internet Crime Report 2020. https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf.

[2] U.S. Department of Justice. Founder of Liberty Reserve Pleads Guilty to Laundering More Than $250 Million through His Digital Currency Business. https://www.justice.gov/opa/pr/founder-liberty-reserve-pleads-guilty-laundering-more\-250-million-through-his-digital.

[3] Agarwal, S. Literature Review on Ponzi Scheme Ecosystem. *COMP0057 Research in Information Security, UCL*, 2021.

[4] Blanton, K. and others. The rise of financial fraud. *Center for Retirement Research Brief*, 2012.

[5] U.S. Securities and Exchange Commission. Ponzi schemes Using virtual Currencies. https://www.sec.gov/files/ia_virtualcurrencies.pdf.

[6] U.S. Securities and Exchange Commission. Ponzi Schemes. https://www.investor.gov/introduction-investing/investing-basics/glossary/ponzi-schemes.

[7] Biography.com Editors. Charles Ponzi Biography. https://www.biography.com/crime-figure/charles-ponzi.

[8] Ponzi Scheme. https://postalmuseum.si.edu/exhibition/behind-the-badge-case-histories-scams-and-schemes/ponzi-scheme.

[9] U.S. Securities and Exchange Commission. Ponzi Schemes. https://www.investor.gov/protect-your-investments/fraud/types-fraud/ponzi-scheme.

[10] Darby, M. In Ponzi We Trust. https://www.smithsonianmag.com/history/in-ponzi-we-trust-64016168/, Dec 1998.

[11] Wikipedia contributors. MMM (Ponzi scheme company). https://en.wikipedia.org/wiki/MMM_(Ponzi_scheme_company).

[12] Sinelschikova, Y. MMM: The biggest fraud in the history of modern Russia. https://www.rbth.com/history/332261-mmm-fraud-mavrodi.

[13] Wikipedia contributors. Bernie Madoff. https://en.wikipedia.org/wiki/Bernie_Madoff.

[14] Hayes, A. Bernie Madoff. https://www.investopedia.com/terms/b/bernard-madoff.asp.

[15] The Editors of Encyclopaedia Britannica. Bernie Madoff. https://www.britannica.com/biography/Bernie-Madoff.

[16] Rushe, D. and Helmore, E. Bernie Madoff, financier behind largest Ponzi scheme in history, dies in prison. https://www.theguardian.com/us-news/2021/apr/14/bernie-madoff-dies-prison-ponzi-scheme.

[17] Wikipedia contributors. High-yield investment program. https://en.wikipedia.org/wiki/High-yield_investment_program.

[18] U.S. Securities and Exchange Commission. High-Yield Investment Programs. https://www.investor.gov/introduction-investing/investing-basics/glossary/high-yield-investment-programs.

[19] U.S. Securities and Exchange Commission. Updated Investor Alert: Social Media and Investing - Avoiding Fraud. https://www.sec.gov/oiea/investor-alerts-bulletins/ia_socialmediafraud.html.

[20] Moore, T. and Han, J. and Clayton, R. The Postmodern Ponzi Scheme: Empirical Analysis of High-Yield Investment Programs. In *Financial Cryptography and Data Security*, pages 41–56, 2012.

[21] U.S. Securities and Exchange Commission. SEC Charges Texas Man With Running Bitcoin-Denominated Ponzi Scheme. https://web.archive.org/web/20130805233337/https://www.sec.gov/servlet/Satellite/News/PressRelease/Detail/PressRelease/1370539730583.

[22] U.S. Department of Justice. Texas Man Sentenced For Operating Bitcoin Ponzi Scheme. https://www.justice.gov/usao-sdny/pr/texas-man-sentenced-operating-bitcoin-ponzi-scheme.

[23] Wikipedia contributors. List of Ponzi schemes. https://en.wikipedia.org/wiki/List_of_Ponzi_schemes.

[24] Vasek, M. and Moore, T. There's no free lunch, even using bitcoin: Tracking the popularity and profits of virtual currency scams. In *Financial Cryptography and Data Security*, pages 44–61, 2015.

[25] Bartoletti, M. and Pes, B. and Serusi, S. Data mining for detecting bitcoin ponzi schemes. In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, pages 75–84, 2018.

[26] Boshmaf, Y. and Elvitigala, C. and Al Jawaheri, H. and Wijesekera, P. and Al Sabah, M. Investigating mmm ponzi scheme on bitcoin. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*, ASIA CCS '20, page 519–530, 2020.

[27] Toyoda, K. and Ohtsuki, T. and Mathiopoulos, P. T. . Identification of high yielding investment programs in bitcoin via transactions pattern analysis. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pages 1–6, 2017.

[28] Toyoda, K. and Ohtsuki, T. and Mathiopoulos, P. T. . Time series analysis for bitcoin transactions: The case of pirate@40's hyip scheme. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 151–155, 2018.

[29] Toyoda, K. and Mathiopoulos, P. T. and Ohtsuki, T. A novel methodology for hyip operators' bitcoin addresses identification. *IEEE Access*, 7:74835–74848, 2019.

[30] Vasek, M. and Moore, T. Analyzing the bitcoin ponzi scheme ecosystem. In *Financial Cryptography and Data Security*, pages 101–112, 2019.

[31] Bartoletti, M. and Carta, S. and Cimoli, T. and Saia, R. Dissecting Ponzi schemes on Ethereum: Identification, analysis, and impact. *Future Generation Computer Systems*, 102:259–277, 2020.

[32] Chen, W. and Zheng, Z. and Ngai, E. C. and Zheng, P. and Zhou, Y. Exploiting Blockchain Data to Detect Smart Ponzi Schemes on Ethereum. *IEEE Access*, 7:37575–37586, 2019.

[33] Chen, W. and Zheng, Z. and Cui, J. and Ngai, E. and Zheng, P. and Zhou, Y. Detecting Ponzi Schemes on Ethereum: Towards Healthier Blockchain Technology. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 1409–1418, 2018.

[34] Neisius, J. and Clayton, R. Orchestrated crime: The high yield investment fraud ecosystem. In *2014 APWG Symposium on Electronic Crime Research (eCrime)*, pages 48–58, 2014.

[35] Drew, J. and Moore, T. Automatic identification of replicated criminal websites using combined clustering. In *2014 IEEE Security and Privacy Workshops*, pages 116–123, 2014.

[36] Wilkins, A. and Acuff, W. and Hermanson, D. Understanding a ponzi scheme: Victims' perspectives. *Journal of Forensic & Investigative Accounting*, 4(1):1–19, 2012.

[37] Amoah, B. Mr ponzi with fraud scheme is knocking: Investors who may open. *Global Business Review*, 19(5):1115–1128, 2018.

[38] M. Chiluwa, I. and Chiluwa, I. "We are a mutual fund:" how Ponzi scheme operators in Nigeria apply indexical markers to shield deception and fraud on their websites, journal = Social Semiotics. 0(0):1–26, 2020.

[39] Xia, P. and Wang, H. and Luo, X. and Wu, L. and Zhou, Y. and Bai, G. and Xu, G. and Huang, G. and Liu, X. Don't Fish in Troubled Waters! Characterizing Coronavirus-themed Cryptocurrency Scams, 2020.

[40] Companies House. Companies House API overview. `https://developer.company-information.service.gov.uk/`.

[41] Companies House. Service information. `http://resources.companieshouse.gov.uk/serviceInformation.shtml`.

[42] Zabor, E. Survival analysis in r. `https://www.emilyzabor.com/tutorials/survival_analysis_in_r_tutorial.html`.

[43] Wikipedia contributors. Censoring (statistics). `https://en.wikipedia.org/wiki/Censoring_(statistics)#Types`.

[44] Kaplan, E. and Meier, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[45] Wikipedia contributors. Logrank test. `https://en.wikipedia.org/wiki/Logrank_test`.

[46] Singh, K. and Gupta, N. Palateless custom bar supported overdenture: A treatment modality to treat patient with severe gag reflex. *Perspectives in Clinical Research*, 2(4):145–148, 2011.

[47] R Cox, D. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[48] MQXB7. MQXB7/COMP0064: Existence of online Ponzi schemes: Insights into new measurements. `https://github.com/MQXB7/COMP0064`, September 2021.

Appendix A

Source Code

In this chapter all the code written to scrape and analyse the High Yield Investment Programs (HYIPs) is provided. Section A.1 displays the python code developed to scrape the aggregators websites. In Section A.2, the Python code snippet to find new HYIPs after scraping aggregators websites is provided. Section A.3, Section A.4 and Section A.5 provides the GitHub repository link to the code used to crawl the individual identified HYIPs, Companies House and GetAddress.io APIs along with some explanations. Finally the analysis code in R is provided in Section A.6, Section A.7, Section A.8 and Section A.9.

A.1 Automatically crawl the aggregators' websites

This section provides the code for crawling the aggregators websites using Selenium, analysing the page source at run time with the help of BeautifulSoup and collect all relevant information like HYIP names in a CSV file.

```
1 from selenium import webdriver
2 from selenium.webdriver.firefox.options import Options
3 from selenium.webdriver.common.desired_capabilities import
      DesiredCapabilities
4 from selenium.common.exceptions import TimeoutException
5 import random
6 from bs4 import BeautifulSoup
7 import re
8 import time
9 import csv
```

```
10 import os
11
12 # define important variables like the initial url to scrape, path
       to geckodriver, etc.
13 DRIVER_PATH = '<Directory Path to geckodriver>'
14 wd = '<Directory Path>/data/main/'
15 wd1 = '<Directory Path>/data/hyips/'
16 wd2 = '<Directory Path>/data/hyiprank/'
17 secondsSinceEpoch = time.time()
18 filename = '<Directory Path>/hyipnames'+ str(secondsSinceEpoch)
       +'.csv'
19 header = (''hyip url")
20 #hyip_url = ''https://hyip.com"
21 hyip_url = ''https://moneymakergroup.com/"
22 total_pages = 1
23 suburls=set()
24 suburls_hyiprank=set()
25 hyiprank_url = ''https://hyiprank.com/details/"
26 hyip_urls=set()
27
28 #define a list of user agents and then rotate them.
29 user_agent_list = [
30     'Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:82.0) Gecko
          /20100101 Firefox/82.0',
31     'Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:82.0) Gecko
          /20100101 Firefox/82.0',
32     'Mozilla/5.0 (X11; Linux i686; rv:82.0) Gecko/20100101 Firefox
          /82.0',
33     'Mozilla/5.0 (Linux x86_64; rv:82.0) Gecko/20100101 Firefox
          /82.0',
34     'Mozilla/5.0 (X11; Ubuntu; Linux i686; rv:82.0) Gecko/20100101
           Firefox/82.0',
35     'Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:82.0) Gecko
          /20100101 Firefox/82.0',
36     'Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:82.0) Gecko
          /20100101 Firefox/82.0',
```

```
37      'Mozilla/5.0 (iPad; CPU OS 10_15_7 like Mac OS X) AppleWebKit
            /605.1.15 (KHTML, like Gecko) FxiOS/29.0 Mobile/15E148
            Safari/605.1.15',
38  ]
39
40  user_agent = random.choice(user_agent_list)
41  #print(user_agent)
42  # Alternate method:
43  #from fake_useragent import UserAgent
44  #useragent = UserAgent()
45
46  def writer(header, data, filename):
47      with open (filename, ''w", newline = '') as csvfile:
48          hyips = csv.writer(csvfile)
49          hyips.writerow([header])
50          for x in data:
51              hyips.writerow([x])
52
53  def find_hyip(source):
54      title=source.find(''meta", property=''og:title")
55      hyip_urls.add(title[''content"].split()[0])
56
57  def page_source(source):
58      #finding all the suburls on a particular page
59      for a in source.find_all('a', href=True):
60       if re.search('/threads/',a['href']):
61           sub=a['href']
62           suburls.add(sub.split('/')[2])
63
64  def hyip_crawl(source):
65      #find all hyip threads on hyip rank and crawl them
66      for a in source.find_all('a', href=True):
67          if re.search('/details/',a['href']):
68              sub=a['href']
69              suburls_hyiprank.add(sub.split('/')[2])
70
```

```
71 def crawl(url, main_page):
72     #use selenium in headless mode
73     options = Options()
74     options.add_argument('--headless')
75     options.add_argument('--no-sandbox')
76     options.add_argument('--ignore-certificate-errors')
77     profile = webdriver.FirefoxProfile()
78     profile.set_preference(''general.useragent.override",
           user_agent)
79     desired_capabilities = DesiredCapabilities.FIREFOX.copy()
80     desired_capabilities['acceptSslCerts'] = True
81     desired_capabilities['acceptInsecureCerts'] = True
82     driver = webdriver.Firefox(firefox_profile=profile,
           desired_capabilities = desired_capabilities, options=
           options, executable_path='<Directory Path to geckodriver>')
83     driver.get(url)
84     driver.implicitly_wait(1000)
85     soup = BeautifulSoup(driver.page_source, 'lxml')
86     secondsSinceEpoch = time.time()
87     if main_page is True:
88      with open(wd + str(driver.title[0:12]) + str(
            secondsSinceEpoch) + '.html', 'w') as f:
89       f.write(driver.page_source)
90     else:
91      check = os.path.isdir(wd1+str(driver.title[0:12]))
92      if not check:
93       os.makedirs(wd1+str(driver.title[0:12]))
94      with open(wd1 + str(driver.title[0:12]) + '/' + str(driver.
            title[0:12]) + str(secondsSinceEpoch) + '.html', 'w') as f
            :
95       f.write(driver.page_source)
96     driver.quit()
97     if soup is not None:
98      soup.prettify()
99      return soup
100
```

```
101 def crawl1(url):
102     #use selenium in headless mode
103     options = Options()
104     options.add_argument('--headless')
105     options.add_argument('--no-sandbox')
106     options.add_argument('--ignore-certificate-errors')
107     profile = webdriver.FirefoxProfile()
108     profile.set_preference(''general.useragent.override",
            user_agent)
109     desired_capabilities = DesiredCapabilities.FIREFOX.copy()
110     desired_capabilities['acceptSslCerts'] = True
111     desired_capabilities['acceptInsecureCerts'] = True
112     driver = webdriver.Firefox(firefox_profile=profile,
            desired_capabilities = desired_capabilities, options=
            options, executable_path='<Directory Path to geckodriver>')
113     driver.get(url)
114     driver.implicitly_wait(1000)
115     soup = BeautifulSoup(driver.page_source, 'lxml')
116     secondsSinceEpoch = time.time()
117     check = os.path.isdir(wd2+str(driver.title[0:12]))
118     if not check:
119         os.makedirs(wd2+str(driver.title[0:12]))
120     with open(wd2 + str(driver.title[0:12]) + '/' + str(driver.
            title[0:12]) + str(secondsSinceEpoch) + '.html', 'w') as f:
121      f.write(driver.page_source)
122     driver.quit()
123     if soup is not None:
124      soup.prettify()
125      return soup
126
127 # to get the main listed hyips from hyip.com/forums/review
128 page = crawl(hyip_url+'/forums/review/', True)
129 page_source(page)
130
131 #finding the total number of pages on the review forum
132 for li in page.findAll(''li", {''class": ''pageNav-page"}): pass
```

```
133 if li:
134     total_pages = li.text
135
136 #recursively parse all the pages on the forum and get all the
        suburls.
137 for i in range(2,int(total_pages)+1):
138     time.sleep(100)
139     page = crawl(hyip_url+'/forums/review/page-'+str(i), True)
140     page_source(page)
141
142 #print names of all hyips and crawl all the hyip threads
143 for i in suburls:
144     print(i)
145     sub_page = crawl(hyip_url + '/threads/' + str(i), False)
146     find_hyip(sub_page)
147
148 writer(header, hyip_urls, '<Directory Path>/hyip_com_'+ str(
        secondsSinceEpoch) +'.csv')
149
150 #crawl hyiprank seperately and get the threads from there directly
        .
151 page = crawl(hyiprank_url, True)
152 hyip_crawl(page)
153 writer(header, suburls_hyiprank, filename)
154
155 for i in suburls_hyiprank:
156     sub_page_hyip = crawl1(hyiprank_url + str(i))
```

A.2 Find new HYIPs after automatic crawling

Crawling the aggregators every alternate day gives us a list of HYIPs through Section A.1 but we need new HYIPs so we can investigate the new ones and not the whole list every time. To do this, we developed a Python code which is available in our GitHub repository at `https://github.com/MQXB7/COMP0064/blob/main/code/check_new_hyip.py` [48]. *We provide a code snippets below for comparing the CSV files of HYIP names as follows:*

```
def compare_csv(filename):
 with open(filename, 'r') as csv_file:
  csv_reader = csv.reader(csv_file, delimiter=',')
  next(csv_reader,None)
  for row in csv_reader:
   if ((str)(row[0])) is not '':
    if ((str)(row[0])) not in hyip_list:
     new_hyip.add((str)(row[0]))
```

A.3 Automatically crawl HYIP website

In order to scrape the HYIPs homepage we developed a code in Python that takes input a list of HYIP website in a CSV file and saves the page in HTML format under a folder with the name of the HYIP. The code is available in the GitHub repository: `https://github.com/MQXB7/COMP0064/blob/main/code/hyip_web.py` [48].

A.4 Companies House API

A.4.1 Searching for companies in Companies House

The Companies House API provides a functionality which can be used to search *companies registered in the Companies House database. We developed the code provided in the GitHub repository:* `https://github.com/MQXB7/COMP0064/blob/main/code/search_companies_house.py` [48] *in order to use the Companies House API to search for the HYIPs claiming to be registered as a "limited company".*

A.4.2 Collecting data from Companies House Register

The Companies House API provides another functionality which can be used to scrape *companies registered in the Companies House database using their registration number. We developed the code provided in the GitHub repository:* `https://github.com/MQXB7/COMP0064/blob/main/code/companieshouse.py` [48] *in order to use the Companies House API to scrape information provided by the HYIPs to the Companies House.*

A.5 GetAddress.io API

Addresses can be easily faked. In order to check the addresses provided by the HYIPs on their websites and to the companies house, we use getaddress.io's API to collect the addresses using postcodes. We later use these list of addresses to verify the address provided by the HYIPs. The code is available in our GitHub repository at `https://github.com/MQXB7/COMP0064/blob/main/code/getaddress.py` [48].

A.6 R code for survival function

We write the following code in R to get the Kaplan-Meier estimator for the survival function and plot the overall survival function of HYIP lifetimes:

```
1  library(survival)
2  library(survminer)
3  library(lubridate)
4
5  hyips<-read.csv(``<Directory Path>/analysis_hyip.csv")
6  km_fit <- survfit(Surv(Number.of.Days, Censored) ~ 1, data=hyips)
7  xmax<-max(hyips$Number.of.Days)
8
9  plot(km_fit, xlim=c(50,xmax), log=``x", xlab = ``Lifetime (Days)",
       ylab = ``Overall Survival Probability", main = ``Survival
     Function of HYIP lifetimes")
10 legend(leg=c(`Lifetime (Observed)',`Lifetime (Observed) (95% CI)')
     ,x=`topright',lty=c(`solid',`dashed'), cex = 0.8)
11
12 #Fit Litecoin variable with survival function
13 fit_litecoin <- survfit(Surv(Lifetime.In.Days, Censored) ~
     Litecoin, data = hyips)
14
15 #for plotting with logarithmic x-axis
16 plot(fit_litecoin, conf.int = TRUE, log=``x", xlim=c(50,xmax), col
     =c(``blue", "red"), xlab = ``Lifetime (Days)", ylab = ``Overall
      Survival Probability", main = ``Survival Function for Litecoin
     ")
```

```
17 lines(fit_litecoin[2], conf.int=TRUE, col="red")
18 legend(leg=c('Does not accepts Litecoin','Accepts Litecoin','
       Accepts Litecoin (95% CI)'),x='topright',col=c('blue','red','
       red'), lty=c(1,1,2), cex = 0.7)
19
20 #Fit Ethereum variable with survival function
21 fit_ethereum <- survfit(Surv(Lifetime.In.Days, Censored) ~
       Ethereum, data = hyips)
22
23 #for plotting with logarithmic x-axis
24 plot(fit_ethereum, conf.int = TRUE, log=``x", xlim=c(50,xmax), col
       =c(``blue", ``red"), xlab = ``Lifetime (Days)", ylab = ``
       Overall Survival Probability", main = ``Survival Function for
       Ethereum")
25 lines(fit_ethereum[2], conf.int=TRUE, col="red")
26 legend(leg=c('Does not accepts Ethereum','Accepts Ethereum','
       Accepts Ethereum (95% CI)'),x='topright',col=c('blue','red','
       red'), lty=c(1,1,2), cex = 0.7)
27
28 #Fit UK Valid Address variable with survival function
29 uk_hyips_data<-read.csv(``<Directory Path>/lifetime_analysis_
       address_UK.csv")
30 fit_valid_add <- survfit(Surv(Lifetime.In.Days, Censored) ~ Valid.
       Address, data = uk_hyips_data)
31 ggsurvplot(fit_valid_add, pval = TRUE, conf.int = TRUE, linetype =
       ``strata", surv.median.line = ``hv", ggtheme = theme_bw(),xlab
       = ``Lifetime (Days)", ylab = ``Overall Survival Probability",
       main = ``Survival Function for Valid Address in UK",palette = c
       (``#E7B800", ``#2E9FDF"), legend.labs = c(``Not Valid UK
       Address", ``Valid UK Address"))
```

A.7 R code for Log-Rank test comparing survival curves

```
1 surv_diff_ethereum <- survdiff(Surv(Lifetime.In.Days, Censored) ~
       Ethereum, data = hyips)
```

```
2 surv_diff_ethereum
3
4 surv_diff_lite <- survdiff(Surv(Lifetime.In.Days, Censored) ~
    Litecoin, data = hyips)
5 surv_diff_lite
6
7 surv_diff_valid_add <- survdiff(Surv(Lifetime.In.Days, Censored) ~
    Valid.Address, data = uk_hyips_data)
8 surv_diff_valid_add
```

Variable Name	Ethereum	Litecoin	Valid UK Address
Ethereum	1		
Litecoin	0.75	1	
Valid UK Address	-0.02	-0.04	1

Table A.1: Correlations among Ethereum, Litecoin and valid UK address: HYIPs, $N = 366$.

A.8 R code for checking correlations between variables

```
1 library(dplyr)
2 library(Hmisc)
3 uk_hyips_var<-read.csv(``<Directory Path>/UK_hyip_variables.csv")
4 total<- hyips %>% left_join(uk_hyips_var)
5 cor_data<-total[,c(16,17,18)]
6 res2 <- rcorr(as.matrix(cor_data))
7
8 #correlation matrix visualisation
9 library(corrplot)
10 cor_data<-total[,c(1,2,3,4,5,7,9,13,14,15,16,17,18,19,20)]
11 res <- rcorr(as.matrix(cor_data))
12 corrplot(res$r, type = ``upper", order = ``hclust", tl.col = ``
    black", tl.srt = 45)
```

A.9 R code for Cox proportional hazards model

```
#combines all variables for UK HYIPs
total<- left_join(uk_hyips_var,hyips)

#multivariate regression
cox_uk_var<-coxph(Surv(Lifetime.In.Days,Censored) ~ Valid.Address
    + Same.Address + Registered.with.Companies.House + Perfect.
    Money + Payeer + Bitcoin + Litecoin + Ethereum + Telegram +
    Facebook + Youtube + Twitter + Instagram + Goldcoders.license.
    check + Contact.Number, data = total)
summary(cox_uk_var)

cox_uk_nonbinary<-coxph(Surv(Lifetime.In.Days,Censored) ~ Valid.
    Address + Same.Address + Registered.with.Companies.House +
    Goldcoders.license.check + Contact.Number + Social.Media.
    Platforms + Payment.Processors, data = total)
summary(cox_uk_nonbinary)

cox_all<-coxph(Surv(Lifetime.In.Days,Censored) ~ Perfect.Money +
    Payeer + Bitcoin + Litecoin + Ethereum + Telegram + Facebook +
    Youtube + Twitter + Instagram + Goldcoders.license.check +
    Contact.Number, data = hyips)
summary(cox_all)

cox_non_binary<-coxph(Surv(Lifetime.In.Days,Censored) ~ Goldcoders
    .license.check + Contact.Number + Social.Media.Platforms +
    Payment.Processors, data = hyips)
summary(cox_non_binary)
```