# Econometrics of Cyber-Investment

## Analysis of Cryptocurrencies and Equity-Crowdfunding

**Xuerui MA**

Supervisor: Prof. Peter Moffatt

Dr. Alasdair Brown

School of Economics

University of East Anglia

This dissertation is submitted for the degree of

*Doctor of Philosophy*

April 2022

I would like to dedicate this thesis to my grandmother Hui Yanping, who will be missed forever . . .

# Declaration

I hereby declare that this thesis is an original report of my research, has been written by me and has not been submitted for any previous degree. The work is entirely my own work, with the exception of Section 4.3 which is co-authored with Prof. Peter Moffatt (UEA) and Dr. Simon Peters (Manchester). The collaborative contributions have been indicated clearly and acknowledged. Due references have been provided on all supporting literature and resources.

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Xuerui MA

April 2022

# Acknowledgements

This thesis would not have been possible without the support of countless people. First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Peter Moffatt for his continuous support of my Ph.D study and related research, and also for his invaluable supervision, patience and motivation. His immense knowledge and plentiful experience have encouraged me in all times of my academic research and writing this thesis as well as daily life. It is my great pleasure and fortune to meet him and I could not have imagined having a better supervisor and mentor for my Ph.D study.

I also deeply appreciate the support and guidance received by Dr. Alasdair Brown and Dr. Fuyu Yang. I am particularly grateful for the insightful comments and support received through the collaborative work undertaken with Prof. Peter Moffatt and Dr. Simon Peters.

I would like to thank all my Ph.D colleagues for going through this journey together and Wenfeng Zhou for teaching me Python. Very special thanks to my friend Dr. Zhiteng Feng for her encouragement since the application along the way to the end of my Ph.D study as well as my friends Mr.& Mrs.Wang, Dr. Alexis Ioannidis and Sara Lei. It is their kind continuing help and support that have made my study and life in the UK a wonderful time.

Last but not least, I would like to express my deepest gratitude to my parents for letting me see this world broadly on top of their shoulders. Also great love to my siblings and cats for their unfailing support, comfort and hope they had given to me, especially during the hard times. Without that hope, I would not have been to this step.

May a wonderful life with happiness all the way with each of you.

# Abstract

This thesis comprises three empirical studies with the common theme of cyber-investment. The first study investigates how price returns of alternative assets such as cryptocurrencies are affected by investor sentiment. The particular focus is on the Bitcoin market, and the gold market is used as a comparator. Many measures of investor sentiment are obtained from a number of online sources including Google Trends. Principal Component Analysis (PCA) is used to reduce the dimension of the sentiment proxies. We find that the price returns of Bitcoin and gold are both responsive to investor sentiment, but that they respond in very different ways. Both the second and third studies set out to predict the level of success of equity crowdfunding campaigns. We have collected data from UK equity crowdfunding platforms, including a measure of the level of success of each campaign. A key feature of this data is that it is truncated. This is because data is only available on campaigns that are successful. It is well known that ordinary least squares regression leads to inconsistent estimation when data is truncated. Because of this, truncated regression models are applied with the results that the target amount has a negative effect on the level of success, and equity provided by the company has an inverted-U-shaped effect. The third study digs deeper into the data source, which is actually panel data: for each successfully funded company, data is available on a sequence of "pitches". The key contribution of this study is the extension of the truncated regression model to the panel setting; the panel truncated regression estimator introduced in this study is new to the econometrics literature. The third study also extends the realm of the results of the second study, by investigating the effects of the number of team members, and also by including quadratic terms which allow us to consider the characteristics of a pitch that maximise (or minimise) the level of success. Moreover, the study includes a simple form of textual analysis, in which the impact of the presence of certain words in the pitch announcement is investigated. Some words are indeed seen to have an important effect.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

**Abbreviations**

AAII American Association of Individual Investors Survey

AIC Akaike information criterion

APARCH Asymmetric Power ARCH

ARCH Autoregressive Conditional Heteroskedasticity

ASSOC Australian Small Scale Offerings Board

AUM Assets Under Management

BBS Bull-Bear Spread

CBO Congressional Budget Office

CBOE Chicago Board Option Exchange

CTO Chief Technology Officer

ECF Equity Crowdfunding

EIS Enterprise Investment Scheme

EMH Efficient Market Hypothesis

EPU U.S. Economic Policy Uncertainty Index

ETFs Exchange Traded Funds

FCA Financial Conduct Authority

FF Federal Funds

FIAPARCH Fractionally Integrated APARCH

FRED Federal Reserve Economic Data

GARCH Generalized Autoregressive Conditional Heteroskedasticity

IRF Impulse Response Function

IRR Internal Rate of Return

VIX Volatility Index

LR Likelihood Ratio

MLEs Maximum Likelihood Estimates

MSL Maximum Simulated Likelihood

OLS Ordinary Least Square

PaaS Platform as a Service

PCA Principal Component Analysis

SaaS Software as a Service

SEIS Seed Enterprise Investment Scheme

STLFSI St. Louis Fed Financial Stress Index

SVI Google Search Volume Index

TED Treasury-EuroDollar rate

TRM Truncated Regression Model

VAR Vector Autoregression

# Chapter 1

# Introduction

## 1.1 Overview

The overarching theme of this thesis is modern online investment, and we will use the term "cyber-investment" since it relates to the use of computers, the internet and technology. It has become a highly important topic due to the rapid development of technology in recent decades. Also, as a result of modern development, the number of retail investors has increased and the higher demand from investors has diverted attention from traditional investments (e.g. stocks and bonds) to alternative investments.[1] Two particular types of alternative investment, both deriving directly from modern technology, are the focus of this thesis: Cryptocurrencies (Bitcoin as the principal example) and Equity crowdfunding (ECF). We are mainly interested in investigating the factors, such as investors' sentiment and success factors of ECF, that may impact the decision-making behaviour of investors in these two markets. By analysing these factors, we aim to gain insights into the underlying drivers of investor behaviour and their implications for asset price returns and the level of success of equity crowdfunding campaigns.

[1] An alternative investment is a financial asset that is not one of the conventional investment categories like stocks, bonds, and cash.

The purpose of this chapter is to provide some background to the two markets of central focus: cryptocurrency (Section 1.2) and crowdfunding (Section 1.3). Also Section 1.4 provides motivation for this research and also outlines the contributions made.

## 1.2 Cryptocurrencies: Background

Cryptocurrencies are a type of digital currency which is decentralised and not controlled by any government. The history of cryptocurrencies can be traced back to the 1980s when they were called cyber-currencies (Narayanan et al., 2016). In the early 1980s, American computer scientist and cryptographer David Chaum conceived an anonymous cryptographic electronic money called ecash (Chaum, 1983). He then introduced DigiCash in 1995, an early version of cryptocurrencies which is untraceable by the issuing bank, the government, or any third party. However, DigiCash only survived for a decade and failed to persuade banks to embrace its technology during that period.

Cryptocurrencies made a breakthrough in 2008 with the introduction of Bitcoin, which was created by an anonymous programmer or group of programmers under the name Satoshi Nakamoto. From this point, the term "cryptocurrency" entered common parlance.

In October 2008, a paper by Nakamoto (2008)[2] titled "Bitcoin: A Peer-to-Peer Electronic Cash System" launched the concept of a system for creating a digital currency that did not require trust in any third party; this publication also outlined the fundamentals of blockchain and Bitcoin. Nakamoto implemented the Bitcoin software as open-source code and released it in January 2009.[3]

In 2010, the first known commercial transaction using Bitcoin happened when programmer Laszlo Hanyecz bought two Papa John's pizzas (total price $25) for 10,000 Bitcoin.[4]

[2] Here we acknowledge that, intriguingly, "Nakamoto" may be a pseudonym for either a person or for a group of individuals.

[3] *New Yorker* website; Davis (2011).

[4] More details of this story can be found in Hankin (2019) on the Investopedia website. It seems that the 10,000 Bitcoin were first transferred to British man Jeremy Sturdivant, who then paid for the pizzas with $25 cash.

In early 2010, Bitcoin was the only cryptocurrency in the market. At that time, its price was just a few cents. Over the following few years, many new digital currencies entered the market, and their prices rose and fell along with Bitcoin's.

Many investors do not have much confidence in Bitcoin because of its volatility. However, Bitcoin appeared to rise to spectacular heights in late 2017. Overall cryptocurrency market capitalisation reached $820 billion in January 2018 before crashing later that month.[5] Meanwhile, Bitcoin has started to be called the New Gold due to its similarities with gold. Also, the number of Bitcoin users continued to grow. Research produced at the University of Cambridge by Hileman and Rauchs (2017) estimated that in 2017, there were 2.9 to 5.8 million unique users using a cryptocurrency wallet, most of them using Bitcoin.

The value of Bitcoin fell dramatically in the early months of 2018 as prices crashed amid uncertainty, fraud and a lack of belief, among other psychological and technical factors. Prices remained comparatively low for a few years, and these negative factors were compounded by the Covid-19 crisis in 2020.

However, since November 2020, possibly as a result of PayPal's decision to offer service for Bitcoin, Ethereum, Litecoin and Bitcoin cash,[6] cryptocurrencies appear to have taken on a new lease of life, and the Bitcoin price reached a maximum above $60,000 in 2021.

Today in 2022, cryptocurrencies - especially Bitcoin - are highly popular as a topic for investigation by researchers including academic researchers (Zhu et al., 2021; Zargar and Kumar, 2019; Gandal et al., 2018). One reason for the intense research interest is the lack of predictability of the market, caused by factors such as the unstable geo-political climate, and the widely-discussed negative environmental impact of the mining process.

[5] Details from Statista (2021).
[6] *TecChurch* website; Perez (2020).

## 1.3 Crowdfunding: Background

The history of crowdfunding is, perhaps surprisingly, long and rich, extending back to the 18th Century.[7] The earliest records of online crowdfunding related to the arts and music communities. Between 1996 and 1997, British rock band Marillion's fans raised $60,000 through an Internet campaign to sponsor their tour in the US.[8] ArtistShare, where artists could seek funding from their supporters to cover their production costs in exchange for free, early access to the artist's album, song or another piece of art,[9] claimed to be the first crowdfunding platform in 2000 and it was inspired by Marillion's innovative method of financing.

Shortly after, more crowdfunding platforms began to emerge, and the crowdfunding industry has grown consistently each year. The first peer-to-peer lending platform – Zopa – kicked off in 2005 in the UK, followed by the launch of Lending Club and Prosper in the US in 2006. In the same year, the term "crowdfunding" was introduced and has been growing fast since then.

Because of the financial crisis between 2008-2009, investors turned their focus away from the conventional market to the Internet. Two now well-known reward-based crowdfunding platforms, IndieGoGo founded in 2008, and Kickstarter founded in 2009, became the mainstream for crowdfunding investment. In just five years, crowdfunding has grown 1,000% and the number of platforms globally has risen to more than 450.[10]

Equity crowdfunding (ECF), one of the types of crowdfunding, is a new way for small businesses to raise money from the public. The first equity crowdfunding platform was launched in 2007 in Australia by the Australian Small Scale Offerings Board (ASSOB). The idea of allowing everyday investors to participate in startup investing without having to deal with high transaction fees and long waits caught on quickly among small business owners

[7] *HistoryWorkshop* website; Clarke (2018) mentioned a young poet, Alexander Pope, who asked donors to pledge two gold guineas to support his work in exchange for having their names published in his book.

[8] *BBC* website; Masters (2013).

[9] Details can be seen Whiteley and Rambarran (2016).

[10] TheWallStreetJournal (2021).

seeking funding for their ventures. The first equity crowdfunding platform in the UK was CrowdCube, launched in 2011. Shortly afterwards, Seedrs was launched and became the first crowdfunding platform to be regulated by the Financial Conduct Authority (FCA). The resulting regulatory framework has been a further factor underlying the accelerated growth of ECF in the UK.

More recently, equity crowdfunding has experienced remarkable expansion: the annual market value of ECF in the UK has risen steadily from £28 million in 2013 to £549 million in 2020, and from US$ 63 million to US$ 280 million over the same period in Europe (excluding the UK).[11] This represents a roughly 40% annual increase. The sector has also seen a rise in the number of high-profile companies using equity crowdfunding to raise funds, such as Monzo and BrewDog. While the industry is still relatively small compared to traditional fundraising methods, it has the potential to become a major player in the investment landscape.

## 1.4 Motivation and Contribution of this Thesis

The motivation for the thesis is clearly seen in terms of the importance of alternative investments at the present time. Alternative investments have grown rapidly in recent years due to their potential to offer higher returns and to bring further diversification to investors' portfolios (Kräussl et al., 2017). The rapid development of cryptocurrencies and crowdfunding has been receiving more attention from a variety of fields, including economic theory, financial econometrics, and small business economics. Over the same period, the number of investors, especially everyday investors without much experience in financial investment, has increased rapidly as a result of modern technology and the development of social media. As the average living standard of people around the world has risen, more people are enjoying a surplus over and above basic needs, and are obviously looking for places to invest this surplus. With regards to higher education level compared to the past and the development of technology such as the internet, which is making the market of alternative investment

[11] Source: https://www.statista.com/statistics/797673/equity-based-crowdfunding-uk/

broader, more people are being attracted by the prospects of the higher returns promised by alternative investments, in preference to straightforward saving.

The clear importance of alternative investments leads to the principal objective of the thesis: to understand the factors affecting the performance of alternative investments. Such an understanding is of obvious value to investors and is also highly relevant in the context of the development of the global economy. The performance of alternative investments is influenced by a multitude of factors that impact consumer protection, risk management, and financial stability.

The thesis contains a number of contributions.

Chapter 2 investigates how the returns of alternative assets are affected by investor's sentiment, using Bitcoin and gold[12] as examples. We collect sentiment data of four different types: market-based, survey-based, news-based and search-based. Principal Component Analysis (PCA) is used to reduce the dimensionality of the sentiment proxies. The contributions arising from this chapter are as follows. First, while the PCA method has been used before to analyse the effect of investors' sentiment on stock returns (see e.g. Baker and Wurgler (2006)), to the best of our knowledge, this is the first time that the PCA method has been applied to returns on alternative investments such as cryptocurrencies. Second, having obtained the Sentiment Indexes using PCA, we apply VAR models to establish that one particular Sentiment Index has a significant lagged impact on the Bitcoin return. We are interpreting this as a violation of the semi-strong form EMH (Malkiel, 1989) and hence a profit-making opportunity for investors. The result also has significant implications for understanding the role of emotions and perceptions in financial decision-making. Third, using Granger causality tests, we confirm that the direction of causality is from the Sentiment indexes to the returns, and not vice versa. Fourth, by using gold as a comparator asset, we establish that gold returns also depend on sentiment, but in a very different way from Bitcoin returns.

Chapters 3 and 4 are both concerned with crowdfunding. In both chapters, we set out to investigate the determinants of crowdfunding success. This is obviously very useful

[12] Gold is used in Chapter 2 as a comparator with Bitcoin.

for companies who are designing a crowdfunding campaign. It is also useful to investors choosing between investment projects, because if a campaign is highly successful, this may be interpreted as a signal that the funded company will also be successful, and the equity held by the investor is likely to rise in value.

An important issue is what is meant by “determinants of success”. The word “success” can be interpreted in more than one way. It could be taken to mean “how successful” the campaign is, with a higher amount raised implying higher success. Or, it could be taken to mean whether the campaign is, or is not, successful, in terms of meeting the target. Given this ambiguity, it is important to make it clear that when the term “success” is used, it usually means the “level of success”, rather than the binary indicator of success.

In both Chapters 3 and 4, the central contribution is that we deal with the problem of truncation in the data. The problem of truncation arises because, at the time we collected the data from the crowdfunding platforms, only information on successfully-funded campaigns[13] was available. Information on unsuccessful campaigns was excluded from the website. For this reason, the data sets that we are working with are truncated data sets. Many previous researchers have succeeded in obtaining data sets containing both successful and unsuccessful campaigns (see e.g. Ahlers et al. (2015). However, it appears that the policy of certain crowdfunding platforms (e.g. Seedrs, Crowdcube) may have changed recently, to the effect that only successful campaigns now appear on the website. We therefore anticipate that the problem of truncated data is likely to become more common in future research, and therefore the methods we use for dealing with truncated data are likely to be called on by other researchers.

In Chapter 3, we analyse a cross-section data set (i.e. with one campaign for each company) and the estimation method we use is the truncated regression model (Hausman and Wise, 1977). This is an established estimation method, but, to the best of our knowledge, this is the first time it has been applied to crowdfunding data. In Chapter 4, we analyse a panel data set. The data is panel data because a sample of companies is observed, and some of the companies

[13] Successfully-funded campaigns are campaigns for which the amount raised is greater than or equal to the target amount.

are observed on more than one occasion (in a sequence of “pitches”). A major contribution of the thesis is that we develop a panel version of the truncated regression estimator (the random effects truncated regression estimator). To the best of our knowledge, this estimator has not been used before and is therefore a contribution to econometric methodology. The importance of using the truncated estimator is confirmed when we apply the Hausman Test (Hausman, 1978), which always shows strong evidence of truncation bias in models that do not allow for truncation.

The importance of using the panel version of the truncated regression estimator is explained as follows. If a pooled estimator was used, it would need to be assumed that all companies are identical. But this is very unlikely. Some companies have a higher propensity to “succeed” in their campaigns than others. That is, there is between-company heterogeneity, which needs to be separated from the equation error. The Random Effects estimator succeeds in separating the error variance into two components: within-company and between-company.

There are other contributions arising from Chapters 3 and 4. We use a range of independent variables to explain the level of crowdfunding success (e.g. target amount, proportion of equity offered, number of team members), and most of these independent variables have been used by other researchers. However, we include quadratic terms in order to allow for non-linear effects. This allows us to deduce the “optimal” level of each independent variable, that is, to identify the value of the variable that maximises the predicted level of success, which is a very useful piece of information for companies needing to raise capital from equity crowdfunding. Also, in one of the models in Chapter 4, we include “word dummies” to investigate the importance of textual content on the success of campaigns. We find that the presence of certain words has a significant effect on the predicted level of success. This is interesting because it provides new information on the sorts of features of campaigns that are likely to attract investors.

# Chapter 2

# Impact of Investor Sentiment on Alternative Investments: Evidence from Bitcoin and Gold

## 2.1 Introduction

Investing money in financial markets has become a popular approach for accumulating wealth. In order to minimise the possibility of financial losses, it is customary to diversify financial portfolios through investment in a range of different asset classes. Alternative investments such as hedge funds, real estate investment trusts and even wine and art are assets not from one of the conventional investment types. They usually have a low correlation with those of standard asset classes in terms of price movements and financial returns. Investors are constantly seeking ways to enhance their portfolio returns while minimising risk exposure, especially after the 2008 financial crisis. To achieve this goal, they have started considering various investment options, including alternative investments offering diversification benefits. Therefore the market for alternative investment is growing and investing in alternative assets has become a popular investment area. According to a report from Preqin (2022), the AUM (Assets Under Management) in the global alternatives industry is expected nearly double to

$18.3tn by the end of 2027. However, the performance of these assets is often affected by a range of factors. Investors' sentiment, which refers to the overall attitude and emotions of investors towards a particular asset or market, can have a significant effect on the prices and returns of financial assets, including alternative investments. Therefore, understanding and monitoring investors' sentiment has become increasingly important for investors who seek to make informed investment decisions and manage their portfolios effectively.

This chapter investigates the relationship between investors' sentiment and the returns of alternative investments, focusing on Bitcoin[1] and gold for comparison. The study uses a statistical technique called Principal Component Analysis (PCA) to construct composite sentiment indices based on variables from market-based, survey-based, news-based, and search-based sentiment. We start off with 11 different sentiment indexes. None of them is particularly interesting when considered individually, and considering all of their effects together would result in a model that is unwieldy and hard to interpret. What is required is a way of collapsing the information into a single sentiment index. The PCA approach is ideal for this purpose.

The main objective of the study is to examine the influence of investors' sentiment on the price returns of alternative assets, because price return is a widely used performance indicator that evaluates the success of investment strategies and the overall performance of financial assets, and price return also forms the basis of many widely-used theories in finance, such as Mean-Variance theory and the Capital Asset Pricing Model (Feibel, 2003). Moreover, Bitcoin, as the dominant cryptocurrency, has been labelled as "New Gold". Therefore, the study analyses the bivariate VAR(2) relationships between each sentiment index and returns for both Bitcoin and gold, firstly to verify that the direction of causality is from sentiment to return, and also to quantify the reaction to investors' sentiment. The results indeed confirm uni-directional causality, and it is found that the sentiment variables relating to economic conditions have a positive impact on the returns of Bitcoin, while sentiment variables relating to the stock market have a negative effect on returns for gold. These findings highlight the

[1] Because Bitcoin is the first created cryptocurrency and is now the most traded (Market Cap around $320 billion by 2023) and well-known (DeVries, 2016).

importance of considering investors' sentiment when investing in alternative assets, as well as the need to differentiate between different types of sentiment indices and their effects on performance. Note also that these results can be used to devise profitable trading rules on the basis of sentiment data, and hence constitute a violation of the semi-strong form Efficient Markets Hypothesis (Malkiel, 1989).[2]

### 2.1.1 Bitcoin

Recently, there has been a growing interest among the general public and academia in cryptocurrencies, particularly Bitcoin, as an alternative asset due to its growing popularity. In May 2010, a programmer named Laszlo Hanyecz purchased two pizzas with 10,000 Bitcoins, considered the first Bitcoin transaction. Today, the highest price of one Bitcoin reached roughly US $65,000 in 2021 and currently exceeds US $20,000. According to a 2017 research study on global cryptocurrency by the University of Cambridge, an estimated 2.9 to 5.8 million unique users hold a cryptocurrency wallet, with the majority using Bitcoin (Hileman and Rauchs, 2017). Furthermore, a recent survey (Sharma et al., 2021) showed that over 70% of respondents from various countries expressed familiarity with Bitcoin and other cryptocurrencies, indicating a significant increase in cryptocurrency adoption.

Despite the proliferation of various forms of cryptocurrency in recent times,[3] Bitcoin holds a distinct place as the first decentralised cryptocurrency to emerge and begin functioning in 2009. Throughout its existence, Bitcoin has experienced a swift and substantial expansion, firmly establishing itself as a noteworthy currency in both the physical and virtual realms. The fundamental feature of cryptocurrency is its decentralised nature, which signifies its independence from the government and central bank authority. According to popular discourse, there exists a peer-to-peer system of electronic cash that facilitates online transactions between parties without the involvement of financial intermediaries (Nakamoto, 2008). Since the middle of the 2010s, several commercial entities have initiated the adoption of Bitcoin

[2] The semi-strong form of the Efficient Markets Hypothesis is that no published information is useful in forecasting stock prices or stock indexes. The sentiment data considered here is an obvious example of published data.

[3] Names of different cryptocurrencies can be found in Appendix.

as a payment mode, in addition to traditional monetary systems. (Chohan, 2017). It has garnered mixed responses from economists, with some expressing approval while others expressing scepticism. For example, Bitcoin has been labelled a speculative economic bubble like the tulip mania of the seventeenth century in Holland (Constancio, 2017), as well as a Ponzi scheme (Braue, 2014). However, the Washington Post pointed out that the cycles of appreciation and depreciation of Bitcoin observed up 5th of November 2013 did not correspond to the definition of speculative bubble (Lee, 2013). Also, Andolfatto (2013), vice president at the Federal Reserve Bank of St. Louis, stated that Bitcoin as an alternate currency poses a potential challenge to the established financial order, thereby offering a valuable mechanism for regulating central banking functions. Moreover, due to their characteristics, Bitcoin developments have drawn the interest of politicians and legislators. Officials in countries such as Brazil, the Isle of Man, Jersey, the United States and the United Kingdom have recognised its ability to provide legitimate financial services. Glaser et al. (2014) argue that most holders of Bitcoin treat Bitcoin as a financial asset allowing investors to diversify their portfolio like a safe haven. As an investment, some Argentinians have bought Bitcoins to protect their savings against the high inflation in their country (Moreno, 2016). Purchases of Bitcoin also increased during the Cypriot financial crisis. More recently, Bitcoin trading in Venezuela has been sky-rocketing amid the 14,000% inflation rate in the country. Local citizens are struggling with the hyper-inflationary crisis, which is making it hard to pay for everyday things, and many are opting to use Bitcoin and other cryptocurrencies. Thus, in countries where the monetary system and financial structures are crumbling, Bitcoin may provide an alternative store of value relative to the local currency (Detrixhe, 2018).

### 2.1.2 Investors' sentiment

Investors' sentiment, which shows the overall attitude or general mood among investors regarding a particular market or asset, has been extensively studied in the equity market in the economy. Additionally, it can also refer to investors' general mood or perspective regarding how the market or a specific investment will perform in the future. This sentiment can be influenced by various factors such as news events, economic indicators, or market trends.

Although it was stated that there is no role for investors' sentiment in classical finance theory (Baker and Wurgler, 2006), according to the existing research, it has been found that some studies, for example, done by Da et al. (2014) and Shu (2010), shown a significant effect of investors' sentiment on asset prices.

Nevertheless, there is a lack of research on the relationship between alternative investments (like Bitcoin) and investors' sentiment. Previous studies' findings about the impact of investors' sentiment on financial assets inspire this paper. After Bitcoin became widely popular in 2013, its price increased from near zero to the historical maximum of roughly US$65,000 in 2021. Academia has started paying more attention to Bitcoin's position in the economy due to its popularity and its prospects to become a global currency (Bukovina et al., 2016). Bitcoin is still relatively new to the economy compared to other investment assets. Also, as the first cryptocurrency in the world and due to its decentralised feature, there has been a notable surge in interest surrounding the factors that influence the value of Bitcoin. Especially according to the examples of countries like Venezuela, where the demand for Bitcoin was high because of hyperinflation and a lack of trust in government policies (Kliber et al., 2019). Also, using Google search queries, one of the sentiment indicators, with specific macroeconomic and financial indicators, Bouoiyour and Selmi (2017) determined that heightened interest in Bitcoin correlates with rising prices in Venezuela. One of the reasons for the conclusion is that people lose confidence in their currency and economy, and this pessimistic sentiment leads to the increased demand for Bitcoin to save their wealth.

### 2.1.3 Bitcoin and Gold

Another alternative investment, gold, is compared with Bitcoin in this chapter because of their similarities. As a physical commodity, gold has been considered a store of value known as a "safe haven" because of its negative correlation with stocks and bonds in terms of their price movements. Cryptocurrencies, particularly Bitcoin, have been called the New Gold in different areas as they share similarities (Dyhrberg, 2016; Klein et al., 2018). Firstly, both assets are "mined" and have limited supply. Gold is difficult to mine and extract, and Bitcoin has a predetermined supply cap of 21 million coins. This limited supply is seen as a key factor in both assets' potential to hold value over time. On top of that, both assets are seen as a means of preserving value that can be used as a hedge against economic uncertainty and inflation. This means they can both potentially provide diversification and protection against market volatility. For millennia, gold has served as a means of storing value owing to its scarcity and intrinsic properties. Some have suggested that Bitcoin is also a possible store of value due to its decentralised nature and predetermined supply (Van Alstyne, 2014). With regard to decentralisation, both gold and Bitcoin are not issued by any particular government or financial institution, which means no central authorities can control their production (Holmes, 2022).[4] However, whether or not gold reacts the same as Bitcoin to investors' sentiment has not been tested. Therefore, it is essential to examine the relationship between investors' sentiment and more than one alternative investment, such as cryptocurrency (Bitcoin) prices and commodities (gold) prices, particularly by comparing the direction and magnitude of the effects. Additionally, it is also important to identify which specific characteristics or types of investors' sentiment have an impact on these price returns. Additionally, it is also important to identify which specific characteristics or types of investors' sentiment have an impact on these price returns.

Therefore, in this study, market-based, survey-based, search-based and news-based sentiment measures are adopted to construct a new sentiment index using the method of Principal Component Analysis (PCA). The new index is then used to investigate the relationship between investor sentiment as a whole and the return of Bitcoin prices. Also, as Bitcoin is treated by some as "digital gold", whether the result would be the same for gold or not would be tested as well for a deeper understanding of the relationship between gold and Bitcoin.

The rest of the chapter is structured as follows: Section 2.2 describes the related empirical literature in financial and cryptocurrency Bitcoin markets as well as literature about investors' sentiment. Section 2.3 describes the data applied in the model, while Section 2.4 contains

[4] Other common currencies, for example, traditional currencies, are not considered because they are often subject to government intervention.

the empirical model in this chapter. Section 2.5 presents the results and also provides a discussion of these results. Finally, Section 2.6 concludes.

## 2.2 Literature Review

### 2.2.1 Bitcoin and Gold

The number of studies about cryptocurrency has grown fast recently, especially for Bitcoin. Some researchers compared Bitcoin with gold to see if they performed the same. By applying data: gold bullion USD/troy ounce rate (Gold Cash), the CMX gold futures 100-ounce rate in USD (Gold Future), the dollar-euro and dollar-pound exchange rates and the Financial Times Stock Exchange Index (FTSE Index) using a GARCH model, Dyhrberg (2016) came up with the result that Bitcoin is similar to gold to most aspects given their comparable hedging capabilities, and comparable symmetric reactions to positive and negative news. However, because of its unique characteristics, Bitcoin is in a position between gold and the dollar, and it can combine some advantages of both commodities and currencies in the financial markets for risk management. Henriques and Sadorsky (2018) compared optimal portfolio weights computed from three different multivariate GARCH models. By including data from five exchange-traded funds (ETFs) in portfolios, they found that portfolios included Bitcoin ranked highest according to risk-adjusted measures. They also show that risk-averse investors are more willing to pay a high-performance fee to switch from a portfolio with gold to a portfolio with Bitcoin. Apart from portfolio-based comparison, Klein et al. (2018) also used APARCH and FIAPARCH models with data from the S&P 500 index, MSCI World and the MSCI Emerging Markets 50 index to compare the properties of Bitcoin and Gold to other markets and assets. Mentioning a limited sample size, they concluded that Bitcoin behaves utterly differently from gold regarding its hedge performances in a portfolio where Bitcoin does not function as a hedge against investments in equities.

### 2.2.2 Investors' Sentiment

Investor sentiment in the economy has been studied extensively in the market literature. One of the well-known studies about investors' sentiment effect is by Baker and Wurgler (2006). Their study is about the argument of how the stock returns would be affected by sentiment. It was specifically stated that classical finance theory leaves no role in investor sentiment. The reason for this is that the semi-strong form of the efficient market hypothesis (Malkiel, 1989) assumes that all published information, including information regarding investor sentiment, has already been factored into stock prices. In Baker and Wurgler (2006), with the application of principal components analysis, a composite sentiment index, which is widely used in other papers about sentiment analysis, was generated based on the first principal component of six following sentiment variables: trading volume of closed-end funds, the first-day returns of initial public offerings (IPOs), the dividend premium, the closed-end fund discount, the number of IPOs, and the equity share in new issues. The study concludes that sentiment does affect the stock return or expected return under specific conditions related to several firm characteristics, including market capitalisation, book-to-market ratio, and past stock returns. Baker and Wurgler (2006) also suggested that a better understanding of sentiment may help with asset pricing.

Apart from the analysis between sentiment and stock return, there are studies about the relationship between sentiment and other financial assets that typically focus on the price, returns, or volatility, such as oil price and option prices (Han, 2007). One is demonstrated that behavioural factors have the power to predict oil price movement (Qadan and Nama, 2018), which means sentiment has a significant effect on oil prices and their volatility. In this study, monthly, weekly and daily data from nine variables were collected to provide a more comprehensive understanding of how investor sentiment affects the price of oil over different time horizons. The Vector Autoregressive (VAR) model was used to analyze the dynamic relationship between investor sentiment and the price of oil by incorporating variables such as the monthly sentiment index mentioned in Baker and Wurgler (2006), the Financial Stress Index published by the Federal Reserve Bank of St. Louis, and the Google search volume from Google Trends websites. It is worth mentioning that Qadan and Nama (2018) found that strong oil market movements attract investors' attention to the oil market. An increase in this attention is followed by greater volatility in the price of oil.

Speaking to search volume, another study (Da et al., 2014) also used daily internet search volume from millions of households to reveal the market-level sentiment. They constructed a Financial and Economic Attitudes Revealed by Search (FEARS) index as a new measurement of investor sentiment. The different part of this study from those above is that the words searched by consumers online were analysed by using Harvard and Lasswell dictionaries, which place words into various categories such as "positive", "negative", "weak", "strong" and so on. After selecting all the online research using these dictionaries, it was easier for them to analyse investors' positive and negative sentiments. Lexicon-based techniques determine the orientation of a document by evaluating the words written against a sentiment or subjective lexicon (Karalevicius et al., 2018). Thus, this kind of lexicon-based sentiment analysis is applied in quite a number of studies about sentiment, especially to those new unconventional assets like Bitcoin. One study also used this technique by Karalevicius et al. (2018) found that interaction between media sentiment and the Bitcoin price exists, and there is a tendency for investors to overreact to news in a short period of time. In this paper, Bitcoin expert media articles were collected by scraping several articles from various Bitcoin-related news portals, and parsing was done using Beautiful Soup, a Python package for parsing HTML and XML documents. Natural language processing techniques, which identify the string of words representing sentiment like adjectives and adverbs in sentence structure (Bukovina et al., 2016), are also used in some other papers to analyse the data. Georgoula et al. (2015) employ the sentiment of the Twitter feed. They collected and analysed a set of tweets during a time period by using keywords related to Bitcoin with Python and MySQL. By using the Vector Autoregressive (VAR) model to examine the relationship between Bitcoin prices and other macroeconomic variables, including the US Dollar Index, gold prices, oil prices, and the SP 500 index, Georgoula et al. (2015) identify that sentiment is considered as a price determinant of Bitcoin in their study.

Furthermore, another paper particularly examines sentiment as a driver of Bitcoin volatility. This paper fills the gap of Georgoula et al. (2015)'s paper and contributes with economic rationale about a link between sentiment and Bitcoin in 2016 by Bukovina et al. (2016). The paper offers a novel approach to decomposing the Bitcoin value between rational and less rational components. Data used in this study is also via Natural language processing techniques from the website Sentdex.com, and the primary source is from the website reddit.com. As for the comparison between gold and Bitcoin, Dyhrberg (2016) found that despite some similarities between Bitcoin, currencies, and commodities, Bitcoin exhibits unique characteristics that position it somewhere between a currency and a commodity as it contains properties of both. Additionally, the study suggests that Bitcoin is highly responsive to market sentiment and can react quickly to investor sentiment changes.

According to all these previous studies, the impact of investors' sentiment has been analysed using various methods and across different assets. The number of studies about Bitcoin also increased a lot recently. Nevertheless, as mentioned in the introduction 2.1, there are not sufficient studies about whether Bitcoin price is related to investor sentiment. Apart from that, there also needs to be more studies about whether investor sentiment affects both Bitcoin and gold price returns, as well as the comparison between them.

## 2.3 Data Description

As mentioned in the beginning, data used to construct a new investors' sentiment in this study are categorised into four, which are market-based, survey-based, news-based and search-based. As Bitcoin is still a new concept compared to conventional assets and it has only been known widely in recent years, the data adopted in this study about those sentiment variables are all from July 2013 to July 2018. Also, since most of them are only weekly data,[5] so all data were selected on a weekly basis for consistency.

[5] A few of the variables in my study do not have data available more frequently than on a weekly basis. Therefore, having all variables on a weekly basis is necessary for consistency and to ensure that all variables are included in the same frequency.

Most of the variables are market-based indices applied in this study. The first one is the Volatility Index ($VIX$) created by the Chicago Board Option Exchange (CBOE), which is also referred to as the "investor fear gauge". It was introduced in 1993 and is computed on a real-time basis throughout each trading day. The $VIX$ replaced the older $VXO$, which was a measure of implied volatility calculated using the 30-day S& P 100 index as the preferred volatility index used by the media (Whaley, 2008). The current $VIX$ (Figure 2.1) index is quoted in percentage points constructed using the implied volatilities on S& P 500 index options and is calculated from options-based theory and current options-market data. It represents the expected range of movement in the S& P 500 index over the following 30 days. For example, if the $VIX$ is 15, this means an expected annualised change with 68% probability (i.e. one standard deviation of the normal probability curve) of less than 15% up or down. An index below 20 generally indicates a stable and stress-free period in the markets, however, Figure 2.1 shows the market was volatile in 2015, where one of the reasons might be the Chinese stock market crash and general emerging market weakness. The same applies to the beginning of the year 2018 during the big market selloff, where the index spiked. Therefore, $VIX$ is a sentiment indicator that can serve as a gauge of market sentiment and can be used to identify periods of heightened fear or uncertainty. Data for the Volatility Index is available from January 1990 and obtained from the Federal Reserve Bank of St. Louis.

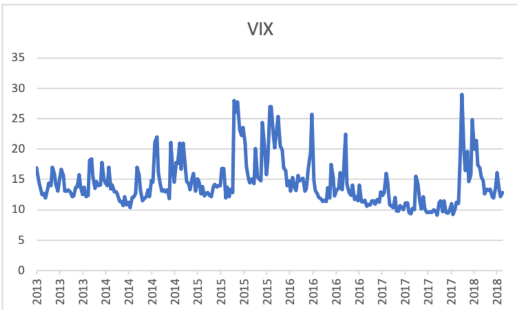



Fig. 2.1 VIX Index from 2013 - 2018

A number of market-based indices are selected from the components used to construct the Financial Stress Index ($STLFSI$), which shows the degree of financial stress in the market. The data has been published on a monthly basis by the Federal Reserve Bank of St.Louis from 1994 to the present. This stress index is calculated based on eighteen weekly data series, which are seven interest rate series, six yield spreads and five other financial indicators. As in this study, $VIX$, which is also a part of $STLFSI$, has also been used as one variable in the model; it would be inaccurate if use $STLFSI$ again due to repetition. Then instead, a few variables used to form STLFSI are adapted to this chapter. They are interest rates of TED spread($\Delta TED_t$), 2-Year Treasury($\Delta TY2R_t$), 10-Year Treasury($\Delta T10YR_t$) and 30-Year Treasury($\Delta T30YR_t$), Effective Federal Funds($\Delta FF_t$),[6] 10-Year Breakeven Inflation Rate($\Delta BI10YR_t$) and BAA Corporate Bond Yield($\Delta BAA_t$). Figures showing their movement over the years are shown below. Notice that in 2016, the TED spread spiked briefly in response to concerns about the global economic outlook, particularly related to China's economic slowdown and the stability of its financial system and also concerns about the stability of the European banking system and the potential for a contagion effect on the global financial system. Furthermore, the Brexit vote in June 2016 added to the uncertainty and volatility in global financial markets, which made the TED Spread spike again.

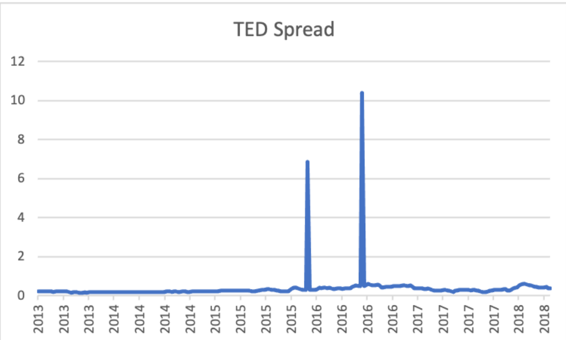



Fig. 2.2 TED Index from 2013 - 2018

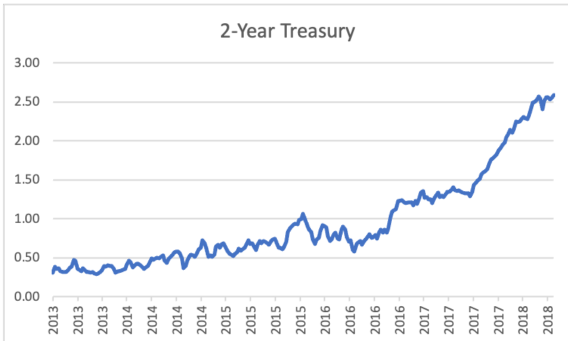



Fig. 2.3 2-Year Treasury from 2013 - 2018

Data on the survey-based sentiment measurement is from the American Association of Individual Investors Survey (**AAII**), which is published on a weekly basis and reported from July 1987 to the present. The association send mails and postcards daily to its members asking their opinions of the stock market and what they think it will do over the next six months. So the survey shows the percentage of individuals who are bullish, bearish and neutral about the stock market over the next six months. The sentiment here is computed as the spread ($BBS$) between the percentage of bullish investors and that of bearish investors

[6] The strange increments pattern of $\Delta FF_t$ is because the weekly of Effective Federal Funds Rate doesn't change often, but when it does, it changes by a large amount.

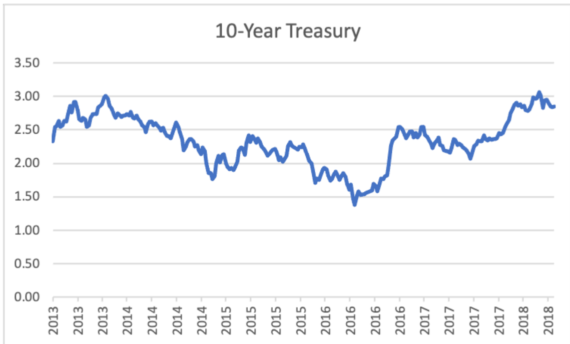



Fig. 2.4 10-Year Treasury from 2013 - 2018

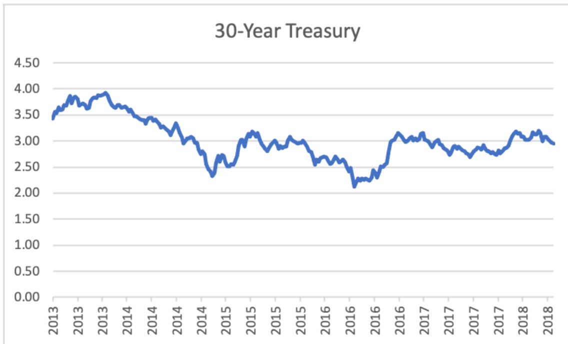



Fig. 2.5 30-Year Treasury from 2013 - 2018

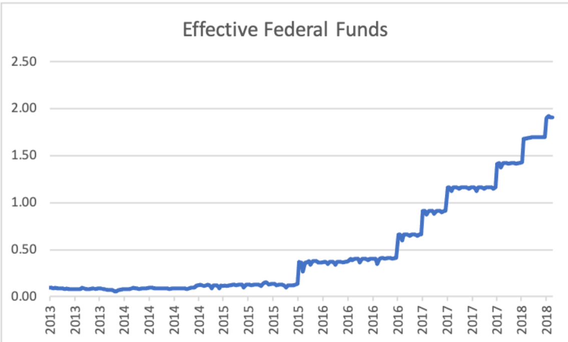



Fig. 2.6 Effective Federal Funds from 2013 - 2018

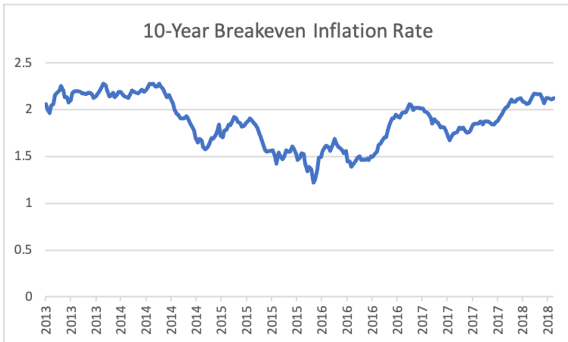



Fig. 2.7 10-Year Breakeven Inflation Rate from 2013 - 2018

(Figure 2.9). For instance, a number of 18 shows there are 18% more members who took the survey and thought the market would be bullish over the next six months.

The news-based sentiment index is data from the U.S. Economic Policy Index ($EPU$) recently developed by Baker et al. (2016) based on newspaper coverage frequency. Three underlying components construct this policy-related economic uncertainty index. The first part is about counting the number of news articles in 10 leading U.S. newspapers that contain specific words such as "economic", "legislation" and so on. The second part is based on reports by The Congressional Budget Office (CBO), which publishes a list of temporary federal tax code provisions. The third part uses the disagreement among the Federal Reserve Bank of Philadelphia's Survey of professional economic forecasters as an indicator of uncertainty. Data here was also collected from the FRED. Figure 2.10 is the plot of $EPU$ from July 2013 to July 2018.

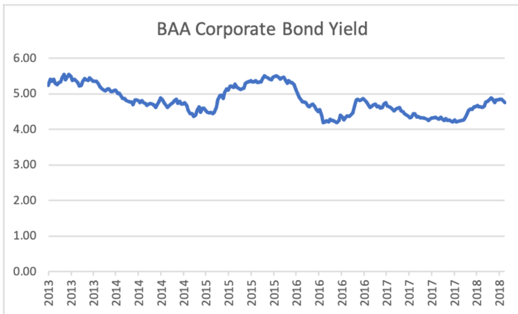



Fig. 2.8 BAA Corporate Bond Yield from 2013 - 2018

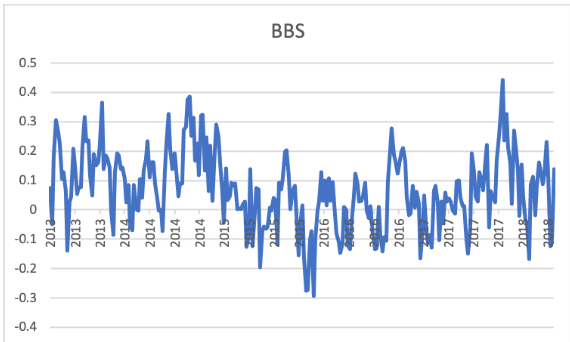



Fig. 2.9 Bull and Bear Spread of **AAII** from 2013 - 2018

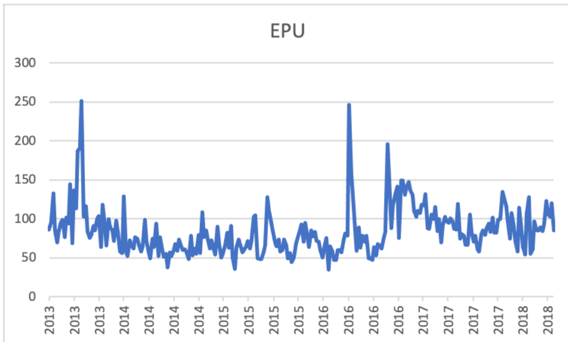



Fig. 2.10 the U.S. Economic Policy Index from 2013 - 2018

The last sentiment measure is the search-based Google Search Volume Index ($SVI$). It shows the degree of individual investor concerns. Weekly data is extracted from the Google Trends website. When a user inputs a specific search term into Google Trend, the website provides the values about search volume history. For their search volumes, "Bitcoin" and "Gold" were put in, respectively. The figures on the chart indicate the level of search interest relative to the highest point within the specified region and time period. A score of 100 reflects the term's highest peak in popularity. This term has recently become a frequently used measure for sentiment because of internet penetration. Qadan and Nama (2018) also mentioned that Google searches mirror some degree of individual investor concerns, at least to some extent, in their research about investor sentiment and oil prices.

The weekly data of the most important variable, Bitcoin prices, were obtained from investing.com. As from Figure 2.13, it varies from the lowest 69.7 US Dollars per Bitcoin to the highest $19,345.5$ US Dollars one Bitcoin during the time period selected. Besides, gold prices were downloaded from the World Gold Council, showing gold prices in various currencies since 1978.

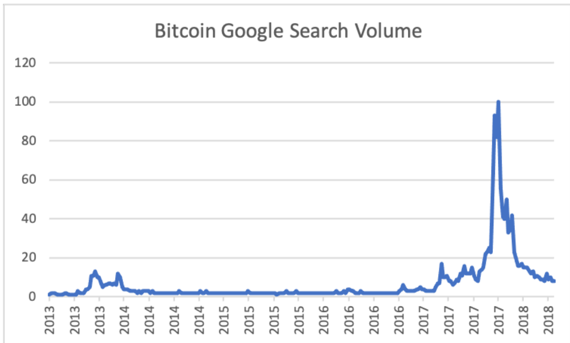



Fig. 2.11 Bitcoin Google Search Volume Index from 2013 - 2018

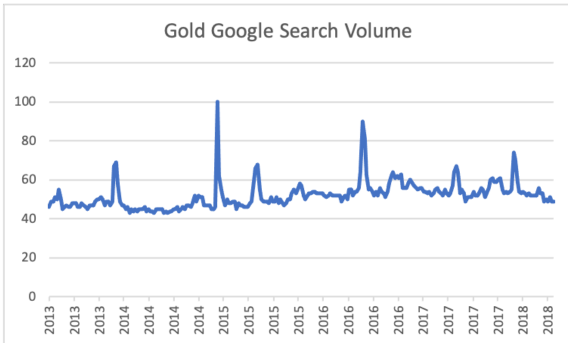



Fig. 2.12 Gold Google Search Volume Index from 2013 - 2018

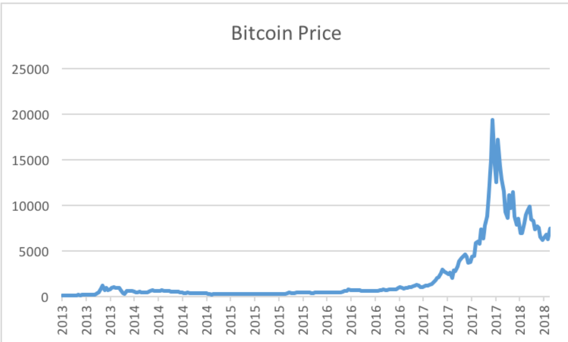



Fig. 2.13 Bitcoin Price 2013 - 2018

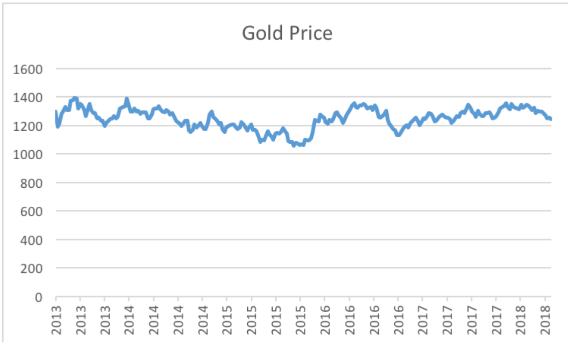



Fig. 2.14 Gold Price from 2013 - 2018

## 2.4 Methodology

The objectives of this study are to construct composite sentiment indices from a few sentiment variables of different areas and to find out whether sentiment factors for testing the influence on conventional investments would also work on alternative investments like Bitcoin and gold. The Principal Component Analysis (PCA) used by Baker and Wurgler (2006) is also adopted here in this research. Principal Component Analysis is a well-established and widely-used technique for analysing large datasets and extracting relevant information from them. It is a popular method in finance and economics for identifying and measuring important factors that contribute to market behaviour. The Baker and Wurgler (2006) study is an important reference point in the literature on investor sentiment, and it was one of the first to apply Principal Component Analysis (PCA) to analyse sentiment in financial markets. Their approach to measuring sentiment has been widely adopted by subsequent studies. (See Chen et al. (2010)). While there may be more recent methodologies available, the Baker and Wurgler (2006) method is still relevant and widely used in the field.

PCA is a multivariate statistical procedure which involves finding the linear combination of a set of possibly correlated variables and reducing the dimensions with a minimum loss of information (Abdi and Williams, 2010). The principal components are obtained by calculating eigenvalues and eigenvectors of the covariance matrix of a vector formed by variables. In the context of principal component analysis (PCA), the eigenvectors represent the directions along which the data exhibits the highest amount of variability. The eigenvalues can be defined as a numerical value that indicates the amount of variance along these directions. Usually, only the principal components with eigenvalues that are greater than 1 are retained according to Kaiser (1960) Criterion because it indicates that the corresponding principal component explains more variance in the data than any single original variable. Therefore, retaining only the principal components with eigenvalues greater than 1 can reduce the dimensionality of the data while retaining most of the variability. On top of these, the first principal component is the linear combination of the original variables in a dataset that captures the maximum amount of variance in the data. It represents the direction of greatest variability in the dataset.

The vector formed by eleven [7] different sentiment variables in this paper is:

$$\mathbf{X}_t =$$

$$\left( \Delta SVI_t \ EPU_t \ VIX_t \ BBS_t \ \Delta TED_t \ \Delta TY2R_t \ \Delta T10YR_t \ \Delta T30YR_t \ \Delta FF_t \ \Delta BI10YR_t \ \Delta BAA_t \right)^T$$

[7] I selected 11 variables that were most relevant to my research question and dataset. This number was also supported by previous literature in the field, which suggests that a larger number of proxies can capture more dimensions of sentiment. Some of the additional variables I included are more recent and relevant to the current market environment than those used by Baker and Wurgler (2006). Additionally, some of the variables used by Baker and Wurgler (2006) were not available for my dataset.

Where $\Delta SVI_t$ is the degree of change of Google Search Volume Index at time t of a specific term like "Bitcoin" when constructing the sentiment index for Bitcoin, and "gold" when constructing the sentiment index for gold, $EPU_t$ is the U.S. Economic Policy Uncertainty Index at time t, $VIX_t$ is the CBOE Volatility Index at time t, $BBS_t$ is the Bull-Bear Spread at time t from the American Association of Individual Investors' Survey and interest rates of TED spread($\Delta TED_t$), 2-Year Treasury($\Delta TY2R_t$), 10-Year Treasury($\Delta T10YR_t$) and 30-Year Treasury($\Delta T30YR_t$), Effective Federal Funds($\Delta FF_t$), 10-Year Breakeven Inflation Rate($\Delta BI10YR_t$) and BAA corporate bond yield($\Delta BAA_t$).[8] All these data have been described in detail in Section 2.3.

The covariance matrix of $\mathbf{X}$ is $\Sigma_{\mathbf{X}_t} = \begin{pmatrix} \sigma^2{}_{\Delta SVI_t} & \cdots & \sigma_{\Delta SVI_t, \Delta BAA_t} \\ \vdots & \ddots & \vdots \\ \sigma_{\Delta BAA_t, \Delta SVI_t} & \cdots & \sigma^2_{\Delta BAA_t} \end{pmatrix}$

The following linear combinations are predicting $I_{i,t}$, $(i = 1,2,3,\cdots 9,10,11)$ from 11 sentiment variables, where $I$ represent the new sentiment index constructed after the application of PCA and $t$ is the time on a weekly basis for all 11 variables:

$$\begin{aligned} I_{1,t} = e_{1,1}\Delta SVI_t + e_{1,2}EPU_t + e_{1,3}VIX_t + e_{1,4}BBS_t + e_{1,5}\Delta TED_t \\ + e_{1,6}\Delta TY2R_t + e_{1,7}\Delta T10YR_t + e_{1,8}\Delta T30YR_t \\ + e_{1,9}\Delta FF_t + e_{1,10}\Delta BI10YR_t + e_{1,11}\Delta BAA_t \end{aligned} \tag{2.1}$$

$$I_{2,t} = e_{2,1}\Delta SVI_t + e_{2,2}EPU_t + e_{2,3}VIX_t + e_{2,4}BBS_t + e_{2,5}\Delta TED_t + e_{2,6}\Delta TY2R_t + e_{2,7}\Delta T10YR_t + e_{2,8}\Delta T30YR_t + e_{2,9}\Delta FF_t + e_{2,10}\Delta BI10YR_t + +e_{2,11}\Delta BAA_t$$

$$\vdots$$

$$I_{11,t} = e_{11,1}\Delta SVI_t + e_{11,2}EPU_t + e_{11,3}VIX_t + e_{11,4}BBS_t + e_{11,5}\Delta TED_t + e_{11,6}\Delta TY2R_t + e_{11,7}\Delta T10YR_t + e_{11,8}\Delta T30YR_t + e_{11,9}\Delta FF_t + e_{11,10}\Delta BI10YR_t + e_{11,11}\Delta BAA_t$$

[8] Details of the ADF test results can be found in Appendix A Table A.2.

The $I_{1,t}$,$I_{2,t}$...$I_{11,t}$ are the all eleven principal components of the analysis. The $e_{i1}$, $e_{i2}$, $e_{i3}$, $e_{i4}$, $e_{i5}$, $e_{i6}$, $e_{i7}$, $e_{i8}$, $e_{i9}$, $e_{i10}$, $e_{i11}$ can be viewed as regression coefficients, where $i = 1,2,3,4,5,6,7,8,9,10,11$.

Collect the coefficients into the vector:

$$\mathbf{e}_i = (e_{i1}, e_{i2}, e_{i3}, e_{i4}, e_{i5}, e_{i6}, e_{i7}, e_{i8}, e_{i9}, e_{i10}, e_{i11})\prime, \text{ where } i = 1,2,3,4,5,6,7,8,9,10,11$$

The variance of $I_{i,t}$ is:

$$var(I_{i,t}) = \mathbf{e}\prime_i \Sigma_{\mathbf{X}_t} \mathbf{e}_i$$

For calculating the coefficients of each component, let $\lambda_1$ through $\lambda_{11}$ denote the eigenvalues of the covariance matrix $\Sigma_{\mathbf{X}_t}$. These are ordered so that $\lambda_1$ has the largest eigenvalue and $\lambda_{11}$ is the smallest. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{11}$.

Let the vector $\mathbf{e}_i$ denote the corresponding eigenvectors $i = 1,2,3,4,5,6,7,8,9,10,11$. It turns out that the elements for these eigenvectors are the coefficients of principal components. The eigenvector $\mathbf{e}_1$ corresponding to the largest eigenvalue $\lambda_1$ is the coefficient vector in $I_{1,t}$, which is the linear combination with maximum variance.

Then the conditions of eigenvalues ($\lambda_i$) and eigenvectors ($\mathbf{e}_i$) of matrix $\Sigma_{\mathbf{X}_t}$ is:

$$\Sigma_{\mathbf{X}_t} \mathbf{e}_i = \lambda_i \mathbf{e}_i$$

Since $\mathbf{e}_i' \Sigma_{\mathbf{X}_t} \mathbf{e}_i$ has no maximum if $\mathbf{e}_i$ is unrestricted, thus we are looking for the maximum of:

$$\lambda_i = \frac{\mathbf{e}_i' \Sigma_{\mathbf{X}_t} \mathbf{e}_i}{\mathbf{e}_i' \mathbf{e}_i}$$

Then:

$$\mathbf{e}\prime_i \Sigma_{\mathbf{X}_t} \mathbf{e}_i = \mathbf{e}_i' \mathbf{e}_i \times \lambda_i$$

$$\mathbf{e}_i' \times (\Sigma_{\mathbf{X}_t} \mathbf{e}_i - \mathbf{e}_i \lambda_i) = 0$$

$$(\Sigma_{\mathbf{X}_t} \mathbf{e}_i - \mathbf{e}_i \lambda_i) = 0$$

So the maximum value of $\lambda_i$ is given by the largest eigenvalue in:

$$(\Sigma_{\mathbf{X}_t} - \lambda_i \mathbf{I}) \mathbf{e}_i = 0$$

Note the $\mathbf{I} \in \mathbb{R}^{11\times 11}$ here is an identity matrix.

As mentioned before, the first principal component is the linear combination of x-variables that has maximum variance (among all linear combinations). It accounts for as much variation in the data as possible. So, $e_{1,1}, e_{1,2}, e_{1,3}, e_{1,4}, e_{1,5}, e_{1,6}, e_{1,7}, e_{1,8}, e_{1,9}, e_{1,10}, e_{1,11}$ are coefficients for the first principal component with:

$$\lambda_1 = var(I_{1,t}) = \mathbf{e}_1' \Sigma_{\mathbf{X}_t} \mathbf{e}_1$$

where (note that this is the first equation in (2.1)):

$$\begin{aligned} I_{1,t} = e_{1,1}\Delta SVI_t + e_{1,2}EPU_t + e_{1,3}VIX_t + e_{1,4}BBS_t + e_{1,5}\Delta TED_t \\ + e_{1,6}\Delta TY2R_t + e_{1,7}\Delta T10YR_t + e_{1,8}\Delta T30YR_t \\ + e_{1,9}\Delta FF_t + e_{1,10}\Delta BI10YR_t + e_{1,11}\Delta BAA_t \end{aligned} \tag{2.2}$$

Therefore, $I_{1,t}$ is the first principal component of vector $\mathbf{X}_t$, and eigenvector $\mathbf{e}_1$ corresponding to $\lambda_1$, is the coefficient of $\mathbf{X}_t$ in $I_{1,t}$.

Apart from that, the bivariate VAR(p) model is used to test the influence of the composite sentiment indices on Bitcoin and gold returns, because a key objective of the research is to identify the direction of causality between price returns and sentiment. Therefore the dependent variables of the VAR models are the price returns of Bitcoin or gold, and the independent variables are the sentiment indices composited by PCA. VAR(2) is chosen because 2 is the optimal lag length of this time series in both Bitcoin and gold models.[9]

The model is written as below :

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{A}_1\mathbf{Y}_{t-1} + \mathbf{A}_2\mathbf{Y}_{t-2} + \varepsilon_t \tag{2.3}$$

Where $\mathbf{Y}_t \in \mathbb{R}^{2\times 1} = (R_t, I_t)'$ which is the vector of prices return of alternative asset Bitcoin or gold($R_t$) (Dependent variable) and Investors Sentiment($I_t$) (Independent variable). $\mathbf{c} \in \mathbb{R}^{2\times 1}$ is the constant vector, $\mathbf{A}_1 \in \mathbb{R}^{2\times 2}$ and $\mathbf{A}_2 \in \mathbb{R}^{2\times 2}$ are two coefficient matrices of corresponding

[9]Details of VAR Lag Order Selection Criteria can be found in Appendix A Table A.3 to Table A.8.

lags, and $\varepsilon_t \in \mathbb{R}^{2\times 1}$ is the vector of error terms satisfying $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon\prime_t) = \Omega$ ($\Omega$ is a $2 \times 2$ positive-semidefinite matrix) and $E(\varepsilon_t \varepsilon'_{t-2}) = 0$.

## 2.5 Empirical Results and Discussion

### 2.5.1 Application of PCA

Before commencing the PCA, the augmented Dickey-Fuller test was applied to all series. The results of these tests are presented in Table A.2 of the Appendix. It was found that there are unit roots in the Bitcoin Price, the Gold Price, the Search Volume Index and the variables from the Stress Index. For this reason, the log-differences of all of these variables are used in the PCA for obtaining the optimal Principal Components. See Patil (2020) for an explanation of the importance of stationarity of variables used in PCA.

A composite sentiment index about Bitcoin ($I_t^B$) is formed by PCA of the covariance matrix constructed by those single sentiment indices. The index is defined as the first principal component of the covariance matrix of variables as equation (2.1). To ensure that the Principal Component Analysis (PCA) results are not impacted by the varying scales and variances of each variable, the index components have been standardised, which is done by subtracting the mean of each variable from its values and dividing by the variable's standard deviation. Hence, all variables are on the same scale and have equal weight in the analysis.[10]

The Bitcoin sentiment index is given by:

$$\begin{aligned} I_t^B = 0.0353\Delta SVI_t^B + 0.0270EPU_t - 0.0968VIX_t + 0.1361BBS_t + 0.0061\Delta TED_t \\ +0.4080\Delta TY2R_t + 0.5377\Delta T10YR_t + 0.5221\Delta T30YR_t \\ -0.0202\Delta FF_t + 0.1219\Delta BI10YR_t + 0.4762\Delta BAA_t \end{aligned} \tag{2.4}$$

From the PCA results, the first principal component explains 30.14% of the sample variance, so it is concluded that one factor captures a large part of the variability in the data. The

[10]Details of each PCA results run by STATA can be found in Appendix Table A.9 to Table A.13.

newly constructed sentiment index here is the linear combination of the original sentiment variables. Signs of coefficient illustrate the direction of each corresponding variable to the interpretation of $I_t^B$ . Thus, only the Volatility index and Effective Federal Fund rate have a negative relationship with the sentiment index $I_t^B$ formed, which means when these two figures increase, investors would be more stressed. Then the sentiment index would decrease (i.e. lower sentiment). According to the PCA results, the absolute values (weight) of Treasury rates as well as BAA corporate bond yield are all around 40% to 50%, which shows they are the most influential variables here in calculating the constructed sentiment index. By comparison, those with small weights, such as TED spread and Effective Federal Fund rate, meaning they are not as influential as the others. Investors feel more secure if the risk-free treasury rate is higher, then the sentiment is high.

Since the Volatility Index and Bull-Bear Spread both indicate the stock market's sentiment,[11] another PCA is done by those sentiment variables only representing economic conditions rather than the conventional market by excluding the variables related to stock market. The result is shown below, where $I_{Et}^B$ is another new composite sentiment index consisting of sentiment variables just showing the whole economic condition. The first component explains 36.17% of the variance, which captures a large part of the variability of data here. Signs of each coefficient are the same as the PCA did before.

The Bitcoin Sentiment Index representing Economic conditions is given by:

$$\begin{aligned} I_{Et}^B &= 0.0368\Delta SVI_t^B + 0.0209EPU_t + 0.0115\Delta TED_t \\ &+0.4082\Delta TY2R_t + 0.5427\Delta T10YR_t + 0.5306\Delta T30YR_t \\ &-0.0162\Delta FF_t + 0.1078\Delta BI10YR_t + 0.4934\Delta BAA_t \end{aligned} \tag{2.5}$$

Therefore, there is another sentiment index $I_{St}$ constructed by PCA as well from those two sentiment variables indicating stock market only. The first principal component captures 60.8% if the variance, which is the most of the variability.

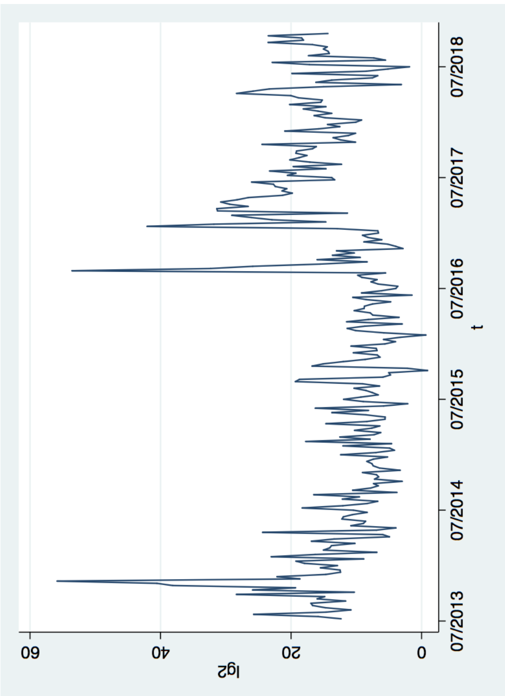
[11] These two variables are firstly collected and generated based on the information from the stock market only, which is different from the other variables.

The Bitcoin Sentiment Index representing the Stock Market is given by:

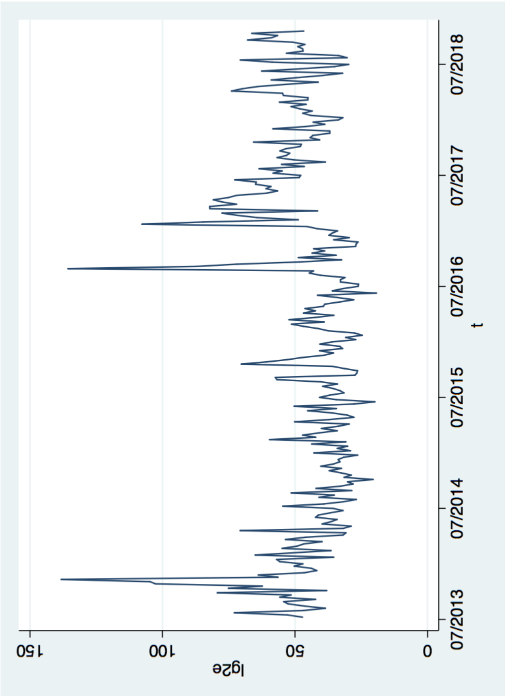
$$I_{S_t} = -0.7071VIX_t + 0.7071BBS_t \tag{2.6}$$

The Index $I_S$ given in Eq (2.6) applies to Gold as well. This is why there is no "B" superscript.

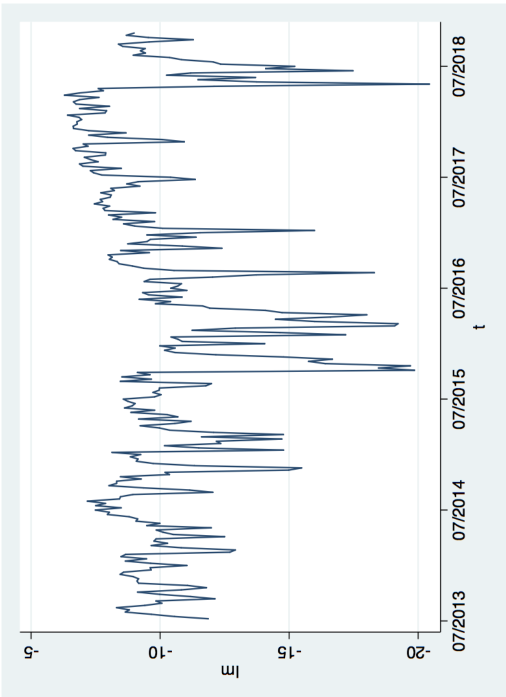
Figure 2.15 shows three plots of the composite sentiment indices of Bitcoin during the time period selected for single sentiment variables. As we can see, the time-paths of $I_t^B$ and $I_{Et}^B$ are similar, which means the investors' sentiment of the whole market did not change much if not considering the sentiment indices from the stock market. However, the movement of the newly constructed sentiment index showing the stock market only varies much more differently.




(a) $I^B_t$ varies with time $t$




(b) $I^B_{Et}$ varies with time $t$




(c) $I_{St}$ varies with time $t$

Fig. 2.15 Sentiment Indices of Bitcoin against time

### 2.5.2 Results of VAR Model and Granger Causality Testing

It is natural to consider the direction of causality between the sentiment indexes and the return variables. We are assuming that the sentiment indexes can be used to predict returns, but we need to check that the direction of causality is not from returns to sentiment. For this, we estimate a sequence of VAR models, each involving one of the sentiment indexes and one of the return variables.

Using the **varsoc** command in STATA, we find that the optimal number of lags (according to the AIC criterion) is 2 for all VAR models (see Tables A3-A8 in Appendix). The results from the three VAR(2) models involving the Bitcoin return are presented in Table 2.1. Whether ${I^B}_t$ or $I^B_{Et}$ is used, we find that the second lag of the index has a significant positive effect on the Bitcoin return. The coefficient of ${I^B}_t$ is +0.0284, meaning that if the sentiment index containing all sentiment variables ${I^B}_t$ increases by one unit, the Bitcoin weekly return is predicted to increase by 0.0284 after a 2-week lag. Similarly, the coefficient of $I^B_{Et}$ is +0.0397, meaning that if the Economic sentiment index $I^B_{Et}$ increases by one unit, the Bitcoin weekly return is predicted to increase by 0.0397 after a 2-week lag.

The results of Granger causality tests are presented in Tables 2.2 - 2.4. These clearly indicate uni-directional (Granger) causality from the (Economic) sentiment index to the price return of Bitcoin.

Figure 2.16 presents Impulse Response Functions (IRFs) from the three VAR models. The top-right plots in Figures 2.16 (a) and (b) both confirm that an upward shock in Economic sentiment is followed after two weeks by an upward jump in the Bitcoin return.

Table 2.1 VAR(2) results of $R_t^B$ and Sentiment Indices

| VARIABLES | $R_t^B$ | $I_t^B$ | $R_t^B$ | $I_{Et}^B$ | $R_t^B$ | $I_{St}$ |
|---|---|---|---|---|---|---|
| L.$R_t^B$ | -0.00818 | -0.138 | -0.00287 | -0.171 | -0.00699 | 0.496 |
| | (0.0606) | (0.330) | (0.0605) | (0.226) | (0.0612) | (0.805) |
| L2.$R_t^B$ | 0.0246 | 0.245 | 0.0303 | 0.103 | 0.0327 | 0.642 |
| | (0.0605) | (0.330) | (0.0606) | (0.226) | (0.0612) | (0.805) |
| L.$I_t^B$ | -0.00487 | 0.490*** | | | | |
| | (0.0112) | (0.0608) | | | | |
| L2.$I_t^B$ | 0.0284** | 0.167*** | | | | |
| | (0.0112) | (0.0612) | | | | |
| L.$I_{Et}^B$ | | | -0.00973 | 0.448*** | | |
| | | | (0.0162) | (0.0605) | | |
| L2.$I_{Et}^B$ | | | 0.0397* | 0.194*** | | |
| | | | (0.0163) | (0.0608) | | |
| L.$I_{St}$ | | | | | 0.00177 | 0.675*** |
| | | | | | (0.00468) | (0.0616) |
| L2.$I_{St}$ | | | | | 0.00362 | 0.0588 |
| | | | | | (0.00467) | (0.0615) |
| Constant | -0.00334 | 0.303*** | -0.0355 | 0.633*** | 0.0721* | -2.715*** |
| | (0.0125) | (0.0683) | (0.0284) | (0.106) | (0.0371) | (0.487) |
| Observations | 263 | | 263 | | 263 | |

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

| Equation | Excluded | chi2 | df | Prob > chi2 |
|---|---|---|---|---|
| $R_t^B$ | $I_t^B$ | 8.0369 | 2 | 0.018 |
| $R_t^B$ | ALL | 8.0369 | 2 | 0.018 |
| $I_t^B$ | $R_t^B$ | 0.72381 | 2 | 0.696 |
| $I_t^B$ | ALL | 0.72381 | 2 | 0.696 |

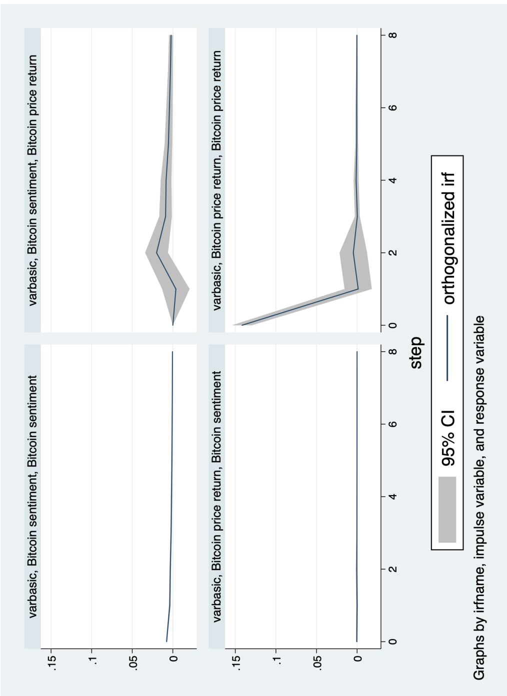
Table 2.2 Granger causality Wald tests ($R_t^B$ & $I_t^B$)

| Equation | Excluded | chi2 | df | Prob > chi2 |
|---|---|---|---|---|
| $R_t^B$ | $I_{Et}^B$ | 7.2599 | 2 | 0.027 |
| $R_t^B$ | ALL | 7.2599 | 2 | 0.027 |
| $I_{Et}^B$ | $R_t^B$ | 0.71501 | 2 | 0.699 |
| $I_{Et}^B$ | ALL | 0.71501 | 2 | 0.699 |

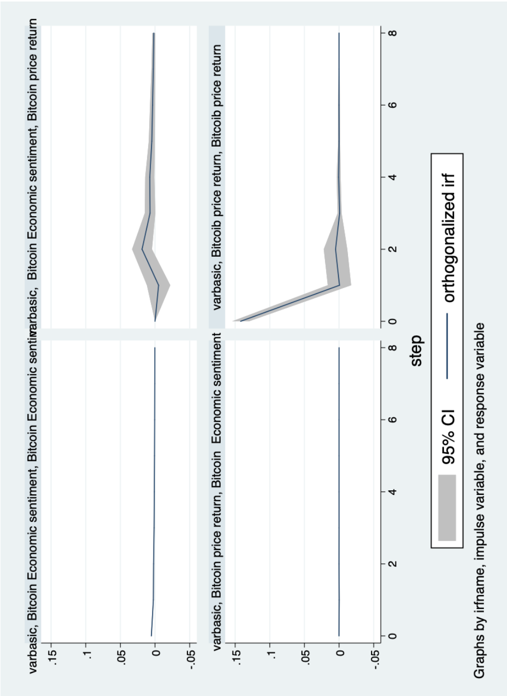
Table 2.3 Granger causality Wald tests ($R_t^B$ & $I_{Et}^B$)

| Equation | Excluded | chi2 | df | Prob > chi2 |
|---|---|---|---|---|
| $R_t^B$ | $I_{St}$ | 2.3954 | 2 | 0.302 |
| $R_t^B$ | ALL | 2.3954 | 2 | 0.302 |
| $I_{St}$ | $R_t^B$ | 1.0169 | 2 | 0.601 |
| $I_{St}$ | ALL | 1.0169 | 2 | 0.601 |

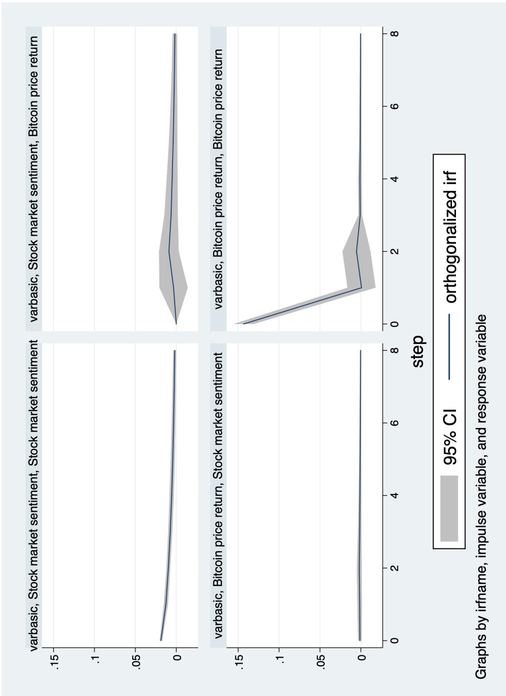
Table 2.4 Granger causality Wald tests ($R_t^B$ & $I_{St}$)




(a) IRFs of Bitcoin sentiment and Bitcoin price return




(b) IRFs of Economic sentiment and Bitcoin price return




(c) IRFs of Stock market sentiment and Bitcoin price return

Fig. 2.16 IRFs of Sentiment Indices and Bitcoin price return

The same procedures are used to test the effect of the constructed sentiment indices on the gold price return. The first principal components leading to $I_t^G$ and $I_{Et}^G$ capture 30.12% and 36.14% of variability respectively. The results are shown below. Remember that the index $I_S$ is the same for Bitcoin and Gold, and this is why there is no "$G$" superscript on $I_S$.

$$\begin{aligned} I_t^G = 0.0196\Delta SVI_t^G + 0.0274EPU_t - 0.0974VIX_t + 0.1370BBS_t + 0.0055\Delta TED_t \\ +0.4081\Delta TY2R_t + 0.5380\Delta T10YR_t + 0.5221\Delta T30YR_t \\ -0.0193\Delta FF_t + 0.1222\Delta BI10YR_t + 0.4761\Delta BAA_t \end{aligned} \tag{2.7}$$

$$\begin{aligned} I_{E\,t}^G = 0.0134\Delta SVI_t^G + 0.0210EPU_t + 0.0111\Delta TED_t \\ +0.4085\Delta TY2R_t + 0.5432\Delta T10YR_t + 0.5308\Delta T30YR_t \\ -0.0156\Delta FF_t + 0.1081\Delta BI10YR_t + 0.4936\Delta BAA_t \end{aligned} \tag{2.8}$$

$$I_{St} = -0.7071VIX_t + 0.7071BBS_t \tag{2.9}$$

The results from the VAR(2) models for Gold are presented in Table 2.5. Only the sentiment index $I_{St}$ has a strong effect on the return of gold prices. It shows that while the first lag of $I_{St}$ (i.e. $I_{S(t-1)}$) increases by one, the return of gold prices would decrease by 0.00185. These results show that even though Bitcoin and gold are compared together due to their similarities. The price returns of gold are only affected by the index showing stock market sentiment only. This is highly likely because gold is always considered by many investors as a "safe haven" asset, while people lose confidence in the stock market, the investors are willing to invest in gold to hedge the risk.

Table 2.5 VAR(2) results of $R_t^G$ and Sentiment Indices

| VARIABLES | $R_t^G$ | $I_t^G$ | $R_t^G$ | $I_{Et}^G$ | $R_t^G$ | $I_{St}$ |
|---|---|---|---|---|---|---|
| L.$R_t^G$ | 0.109* | 1.144 | 0.110* | 0.918 | 0.0997* | 0.77 |
| | (0.0588) | (2.388) | (0.0589) | (1.626) | (0.0584) | (5.797) |
| L2.$R_t^G$ | -0.177*** | 1.782 | -0.179*** | 1.520 | -0.182*** | 0.886 |
| | (0.0576) | (2.336) | (0.0576) | (1.590) | (0.057) | (5.655) |
| L.$I_t^G$ | -0.00178 | 0.487*** | | | | |
| | (0.00150) | (0.0608) | | | | |
| L2.$I_t^G$ | 0.00195 | 0.174*** | | | | |
| | (0.00150) | (0.0609) | | | | |
| L.$I_{Et}^G$ | | | -0.000523 | 0.444*** | | |
| | | | (0.00219) | (0.0604) | | |
| L2.$I_{Et}^G$ | | | 0.00222 | 0.195*** | | |
| | | | (0.00219) | (0.0606) | | |
| L.$I_{St}$ | | | | | -0.00185*** | 0.680*** |
| | | | | | (0.000621) | (0.0616) |
| L2.$I_{St}$ | | | | | 0.00115* | 0.063 |
| | | | | | (0.000622) | (0.0617) |
| Constant | -0.000503 | 0.312*** | -0.00291 | 0.640*** | -0.00695 | -2.606*** |
| | (0.00172) | (0.0698) | (0.00385) | (0.106) | (0.0049) | (0.486) |
| Observations | 263 | | 263 | | 263 | |

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

On top of that, it is clear that there is also uni-directional (Granger) causality from the stock market sentiment index ($I_{St}$) to price returns of gold ($R_t^G$) by the results of the Granger test shown in Table 2.8. This is further confirmed by the IRF's presented in Figure 2.17.

| Equation | Excluded | chi2 | df | Prob > chi2 |
|---|---|---|---|---|
| $R_t^G$ | $I_t^G$ | 1.9673 | 2 | 0.374 |
| $R_t^G$ | ALL | 1.9673 | 2 | 0.374 |
| $I_t^G$ | $R_t^G$ | 0.9301 | 2 | 0.628 |
| $I_t^G$ | ALL | 0.9301 | 2 | 0.628 |

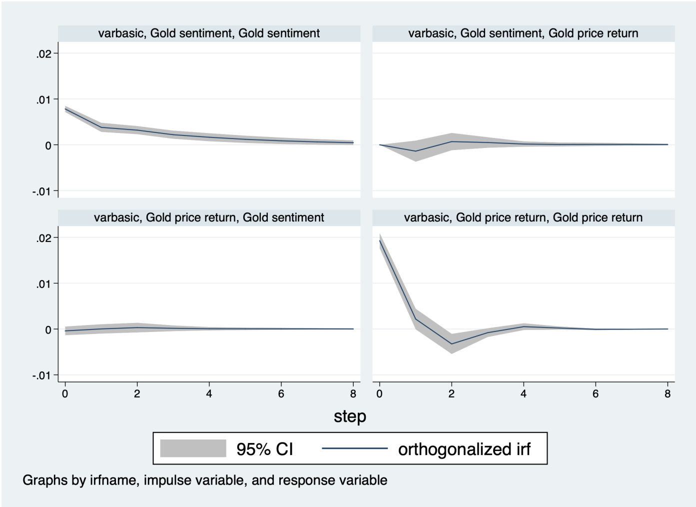
Table 2.6 Granger causality Wald tests ($R_t^G$ & $I_t^G$)

| Equation | Excluded | chi2 | df | Prob > chi2 |
|---|---|---|---|---|
| $R_t^G$ | $I_{E\,t}^G$ | 1.0373 | 2 | 0.595 |
| $R_t^G$ | ALL | 1.0373 | 2 | 0.595 |
| $I_{E\,t}^G$ | $R_t^G$ | 0.72395 | 2 | 0.696 |
| $I_{E\,t}^G$ | ALL | 0.72395 | 2 | 0.696 |

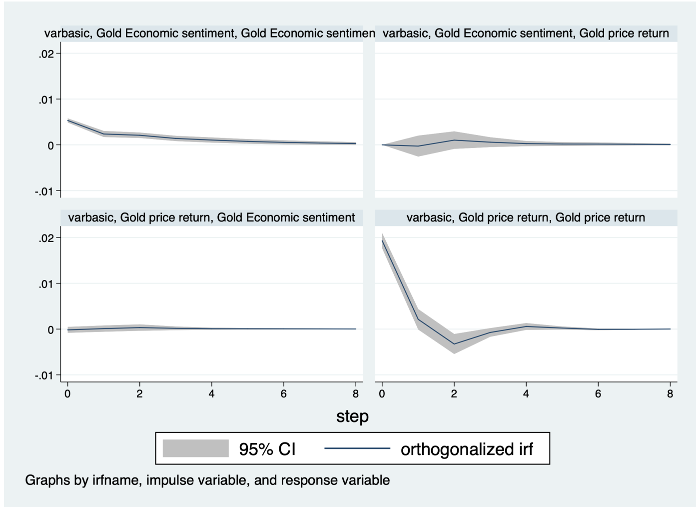
Table 2.7 Granger causality Wald tests ($R_t^G$ & $I_{E\,t}^G$)

| Equation | Excluded | chi2 | df | Prob > chi2 |
|---|---|---|---|---|
| $R_t^G$ | $I_{St}$ | 9.0313 | 2 | 0.009 |
| $R_t^G$ | ALL | 9.0313 | 2 | 0.009 |
| $I_{St}$ | $R_t^G$ | 0.04817 | 2 | 0.976 |
| $I_{St}$ | ALL | 0.04817 | 2 | 0.976 |

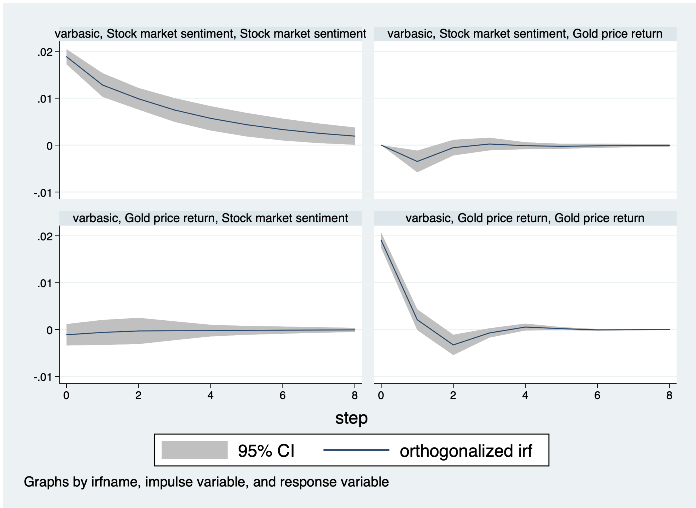
Table 2.8 Granger causality Wald tests ($R_t^G$ & $I_{St}$)




(a) IRFs of Gold Sentiment Index and Gold price return




(b) IRFs of Economic sentiment and Gold price return




(c) IRFs of Stock market sentiment and Gold price return

Fig. 2.17 IRFs of Sentiment Indices and Gold price return

## 2.6 Conclusion

This chapter has focused on the cryptocurrency Bitcoin, and the precious metal gold. One reason for focusing on these two assets is their popularity: Bitcoin is undoubtedly the dominant cryptocurrency, while gold, as a consequence of its intrinsic qualities, plays a very important role in investors' portfolio decisions. Another reason for focusing on them is their similarities to each other. Both may be considered as examples of alternative investments, and it has been suggested that Bitcoin might fulfil the role of the"safe haven", similar to gold. For these reasons, Bitcoin is sometimes referred to as "digital gold".

The chapter has been concerned with the question of whether it is possible to predict returns on these two assets using "investors' sentiment" variables. These sentiment variables are intended to capture the general mood among investors in the whole market. They were collected from four categories: market-based; survey-based; news-based; and search-based.

One of our main objectives was to investigate the influence of investors' sentiment in a broad way rather than to focus on the effect of any specific sentiment index. For this reason, we have used principal component analysis (PCA) to combine a group of indexes into a single sentiment index. The first principal component, indicated by the highest eigenvalue, accounts for the largest proportion of variability in the data. This variable is used as the composite sentiment index.

The original sentiment indexes have been combined in different ways, to give an Index representing Economic conditions relevant to the asset in question, an Index representing stock market conditions, and an overall index.

The relationships between the composite indexes and the return variables were then analysed using VAR(2) models. Key results from these models are: there is unidirectional (Granger) causality from the sentiment variables to the return variables; sentiment relating to economic conditions has a positive lagged impact on Bitcoin returns; sentiment relating to stock market conditions has a negative lagged impact on gold returns.

These results demonstrate that while the returns of Bitcoin and gold are both dependent on investor sentiment, their reactions to investors' sentiment are very different. Therefore, when investors are choosing from alternative investments to diversify the risk of their portfolio, information on investors' sentiment should not be ignored. One useful way of interpreting these results is as a violation of the semi-strong form of the Efficient Markets Hypothesis (EMH), because published information (in the form of sentiment indexes) is useful in predicting returns (Malkiel, 1989). A consequence is that it is possible for investors to apply a "profitable trading rule" to information on investor sentiment in order to generate abnormal profits.

This chapter has potential limitations which can form the basis for future research. For example, the data is not up to date. Importantly, it is possible that dramatic changes have taken place as a result of the Covid-19 pandemic. An obvious possibility for future study is therefore to carry out the analysis of this chapter using up-to-date data. Also, the cryptocurrency market has been rapidly changing, and cryptocurrencies other than Bitcoin may have reached "mature" status. An obvious possibility is therefore to extend the analysis to other popular cryptocurrencies. And of course, it may be useful to consider alternatives to gold as the "comparator" asset. Obvious choices for this role would be major currencies such as the US dollar, the Euro, or the UK pound.

# Chapter 3

# Predicting the Success of Equity Crowdfunding Campaigns: Using Cross-Sectional Data

## 3.1 Introduction

The number of investors has been increasing in recent decades due to various factors such as easier access to investment platforms and increased awareness about investing. Moreover, the internet holds significant potential to drive economic growth and productivity through the transformation of various sectors of the economy, including retail, media, and financial services (Manyika and Roxburgh, 2011). In the past decade, crowdfunding, an online alternative financing tool for venture funding (Mollick, 2014; Kuppuswamy and Bayus, 2017; Fukuhara et al., 2020), has become one of the new trends for gathering capital for start-up companies (Kleemann et al., 2008; Agrawal et al., 2014; Elendner et al., 2017). It can be described as a combination of two familiar concepts: microfinance and crowdsourcing (Bradford, 2012). In crowdsourcing, a 'crowd' or group usually contribute information, ideas or services rather than tangible capital. As for microfinancing, it is about lending very small loans to unemployed, low-income individuals or groups who can not access traditional banking services Yasar (2021). It provides financial services to a large group of people rather than gathering from the crowd. Therefore, crowdfunding, which raises capital for a project or venture from a large number of people, has grown fast and already impacted traditional financing institutions (Khavul et al., 2013). One of the reasons is that small businesses, which are not listed companies, are not able to get enough capital from people they know around them, such as their family and friends. These companies are neither big enough to get investment from banks nor angel investors because of their relatively new innovative business ideas and higher risk as well as low return compared to those conventional investment schemes (Klöhn and Hornuf, 2012). Thus, crowdfunding is treated as a considerable way for small businesses to raise capital at the beginning stage of the company (Hornuf and Schmitt, 2016) or for some of them to expand their business. Another reason for the increasing popularity of crowdfunding, apart from minimising the financing gap, is that it also enhances innovations. It is suggested that crowdfunding can provide access to funding, reduce information asymmetry, enable customer engagement, and create a community of supporters. Innovative projects are necessary for entrepreneurs who seek to raise funding (Hervé and Schwienbacher, 2018). This would promote the development of a diversity of business ideas.

This chapter aims to predict the level of success of equity crowdfunding campaigns. The results obtained are likely to be of direct interest to companies because they convey important information about the ways in which a campaign should be approached in order to maximise the amount raised. The results are also of indirect interest to investors, because it might be assumed that a campaign that is successful in raising funds is also likely to do well when the funded project gets underway.

To achieve this aim, data on successfully funded campaigns is collected from Crowdcube and Seedrs, two prominent UK equity crowdfunding platforms. Given that only successful campaigns can be observed, truncated regression model is adopted to identify the factors that determine the level of success.[1] The comparison of truncated regression results with OLS

[1] Two depedent variables are applied in this chapter to measure the level of success: log of amount raised and log of success ratio. Success ratio is the ratio between the amount raised and the target amount.

results reveals the presence of truncation bias when OLS is employed. Also, the analysis reveals that the target amount of each campaign has a negative impact on its level of success. Additionally, the equity provided by the company has an inverted-U shaped effect on success, where a moderate amount of equity is appealing to investors, but too much equity lowers investors' confidence in the campaign. By estimating the optimal amount of equity for companies to provide, valuable insights are provided for equity crowdfunding campaigns.

### 3.1.1 Crowdfunding

A crowdfunding raising process consists of three parties (Stevenson et al., 2019): The entrepreneur, the start-up company that has the innovation, the crowd, the everyday investor who provides the funding and the online platform that provides the connection between entrepreneurs and investors (Pollack et al., 2021). It has been categorised into four parts: donation, reward, lending and equity crowdfunding (Vulkan et al., 2016). Donation refers to non-government institutions or private parties donating for charity purposes (Block et al., 2018). Reward refers to the physical return that investors receive in exchange for their investment, such as products or services offered by the company in which they have invested (Cox and Nguyen, 2017). Crowdfunding lending, one of the most popular ways of crowdfunding, is also explained by the concept called peer-to-peer(P2P) lending, which transfers the traditional way of loaning into the internet (Bachmann et al., 2011). It is based on the lender (people lending money) receiving an interest rate as a return from the borrower (people borrowing the money) (Herzenstein et al., 2011; Lin et al., 2013). Lastly, equity crowdfunding, the main focus of this chapter and the following chapter, is the process whereby people (i.e. the 'crowd') invest on internet-based platforms in an early-stage private company which is not on the stock market in exchange for shares in that company (Walthoff-Borm et al., 2018). Compared with the other three types of crowdfunding, P2P or marketplace lending (Cumming and Hornuf, 2020) and equity crowdfunding is investment-based crowdfunding. However, people invest in equity crowdfunding to get shares of the company, and they may also have partial ownership of the funded company. On top of that, equity crowdfunding is considered to be the type of crowdfunding that involves the highest risk. This is due to the implication of the risk-return equation, which suggests that higher potential returns are associated with higher levels of risk (Bapna, 2019). Since equity crowdfunding involves an uncertain financial return, unknown future date, and a possibility of losing everything, it is considered to be the crowdfunding with the highest level of risk (Coakley et al., 2021).

### 3.1.2 Equity Crowdfunding

More recently, equity crowdfunding has an increased global use (Habla and Broby, 2019), especially for innovative ventures (Audretsch et al., 2016; Bruton et al., 2015; Vismara, 2021). It has become more important and provides a new opportunity to do research on smaller private businesses for a better understanding of entrepreneurial finance (Coakley et al., 2021; Konhäusner et al., 2021). It is a new way for small firms to raise money from everyday investors. It allows people who are not accredited investors (those with more than $200,000 in annual income or net worth) to invest in start-ups and companies that do business on the Internet. Equity crowdfunding is similar to Kickstarter, one of the largest crowdfunding platforms, but instead of asking people to donate money, companies ask people to invest in their businesses.

Equity crowdfunding works relatively the same way as other types of crowdfunding: Start-up companies who want to raise money from the crowd create a campaign on a website and share their idea with potential investors. When investors decide whether to invest, they buy shares in the company through an online equity crowdfunding platform. If enough people give the company money, it can be used to fund the project or even pay off debt. Furthermore, there is a low entry level of requirements for companies that want to use equity crowdfunding. It is for any business that wants to raise money from investors. This is especially useful for small businesses and start-ups, but it can also be used by established companies with a proven track record of success. From the company's point of view, they created a campaign on the website of an online crowdfunding platform, such as Seedrs or Crowdcube (platforms analysed in this chapter). The company then needs to share their idea with potential investors, who will decide whether they want to invest in the company. If the company gets enough people interested in investing, it can use investors' money to fund the project or pay off debt. If the company has a successful campaign and reaches its funding goal, the amount that the company raises will depend on how many shares are sold. It will be able to see how much each investor has invested and how much they have earned when the project is complete.

Equity crowdfunding platforms have been rapidly appearing on the UK funding scene since 2011 (Bank, 2014). According to Statista (2017), the annual transaction value of equity-based crowdfunding in Europe (excluding the UK) has increased from 18.4 million euros in 2012 to 211 million in 2017. In the UK, the annual equity transaction volume increased from 30 million in 2011 to 333 million in 2017. Previously only wealthy individuals, venture capitalists and business angels could invest in start-ups because of their experience and sufficient wealth. However, equity crowdfunding platforms have helped democratise the investment process by providing more chances to a larger pool of potential investors (InterTradeIreland, 2016; Nutting and Freedman, 2015). Apart from that, the UK government offers tax reliefs on eligible opportunities in the form of the Enterprise Investment Scheme (EIS) and the Seed Enterprise Investment Scheme (SEIS) to offset the partial risk involved with investing in early-stage companies. These are some of the most generous tax incentives in the world, offering 30% to 50% income tax relief, respectively. Therefore, as the requirements of investors are much simpler, individuals who would like to invest their capital in the financial market are becoming more likely to choose equity crowdfunding and to be a shareholder of the company they would like to invest in.

Equity crowdfunding platform Seedrs has published a portfolio update of issuers raising capital on its platform. According to the company's report, the 577 transactions that garnered funds from July 2012 until the conclusion of 2017 have yielded a platform-wide internal rate of return (IRR) of 12.02%. Once the impact of both the EIS and SEIS tax advantages are taken into consideration, the IRR jumps to 26.42%. However, investing in start-up companies is still risky for non-professional investors. Not all crowdfunding campaigns would raise enough amount of funds needed.

Figure 3.1 illustrates the Google Search Volume[2] of the term "Equity Crowdfunding " indicating a noticeable increase in the public's interest in equity crowdfunding. This can be attributed to the fact that equity crowdfunding enables broader participation in investments, in contrast to traditional methods that entail high entry barriers for investors. However, as investors in equity crowdfunding tend to possess lower levels of experience and encounter significant information asymmetries when evaluating investment opportunities (Ahlers et al., 2015; Bapna, 2019), there is a pressing need to examine the determinants of equity crowdfunding success. Such analysis can provide significant and valuable insights for various stakeholders, including entrepreneurs, investors, and policymakers interested in participating in or regulating the rapidly growing field of crowdfunding.

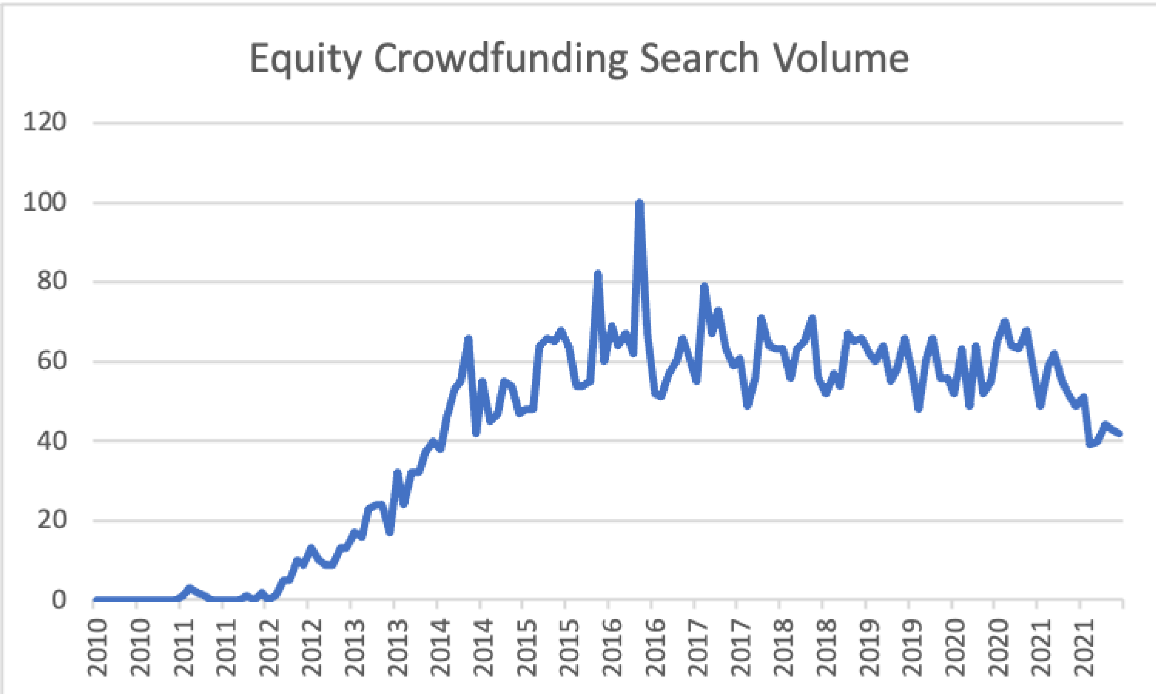



Fig. 3.1 Equity Crowdfunding Google Search Volume

### 3.1.3 Data Truncation

On top of that, incomplete data has been analysed and applied in a lot of different studies (Ghahramani and Jordan, 1995; Hartley and Hocking, 1971) and the problem of incomplete

[2]Details of Google Search Volume is mentioned in Chapter 2 Section 2.3.

data exists in a wide range of research areas (Williams et al., 2007). Considering the characteristics of data collected for this chapter, data truncation should be aware and considered. Truncated data has been compared with censored data by studies (Breen et al., 1996; Honoré and Powell, 1994; Mandel, 2007). Manual (1989) mentioned that truncation and censoring are different concepts. Censored data contain at least partial information for observation outside the given boundary, whereas truncated data does not even observe those outside the given boundary (Klein and Moeschberger, 2003). For instance, Lee et al. (2013) examined the survival of individuals with advanced lung cancer. The study was right-censored since some patients were still alive at the end of the study period, and their survival times were unknown beyond that point. On the other hand, Trussell and Bloom (1979) mentioned that the sample on the height of Royal Marines is truncated because those below the minimum allowed height do not appear in the sample at all as a result of the minimum height restrictions for the recruits. This example clearly emphasises the frequency truncation of data in real life. Furthermore, truncation could not be ignored and pretend the dataset is complete data because the sample average would be inconsistent for the population mean because all observations below the truncation point are missing (Canette, 2016). Therefore, consideration of the truncated data in research becomes important.

In this chapter, to minimise the information asymmetries, we set out to predict the success of the equity crowdfunding campaigns by analysing the data showing those successfully funded campaigns collected from the two large UK equity crowdfunding platforms Crowdcube and Seedrs. Then both potential investors and small companies that would like to raise capital from equity crowdfunding would have a clearer idea of the determinants of campaign success with the consideration of data truncation bias. The chapter also emphasises the importance of truncated data consideration, as it appears a lot in daily life and is easy to be ignored.

The rest of this chapter proceeds as following: Section 3.2 introduces the previous literature and work done by other researchers related to crowdfunding and equity crowdfunding success. Section 3.3 briefly presents the truncated data applied. Section 3.4 shows the main focus;

truncated regression model. Section 3.5 discusses the primary results of the model. Finally, Section 3.6 concludes.

## 3.2 Literature Review

Literature on crowdfunding has grown rapidly in the last decade. According to a research paper by Angerer et al. (2017), the emphasis is on the development of equity crowdfunding in academic research area by showing the number of journal articles based on the search "equity crowdfunding" on the website ProQuest and the search "crowdfunding success" on ScienceDirect. Figure 3.2 shows an extension (from 2015 until the end of 2021) of what they did, and the number of articles increased significantly from less than 50 each year to over 300 more recently. It demonstrates the increase of research interests from an academic aspect.

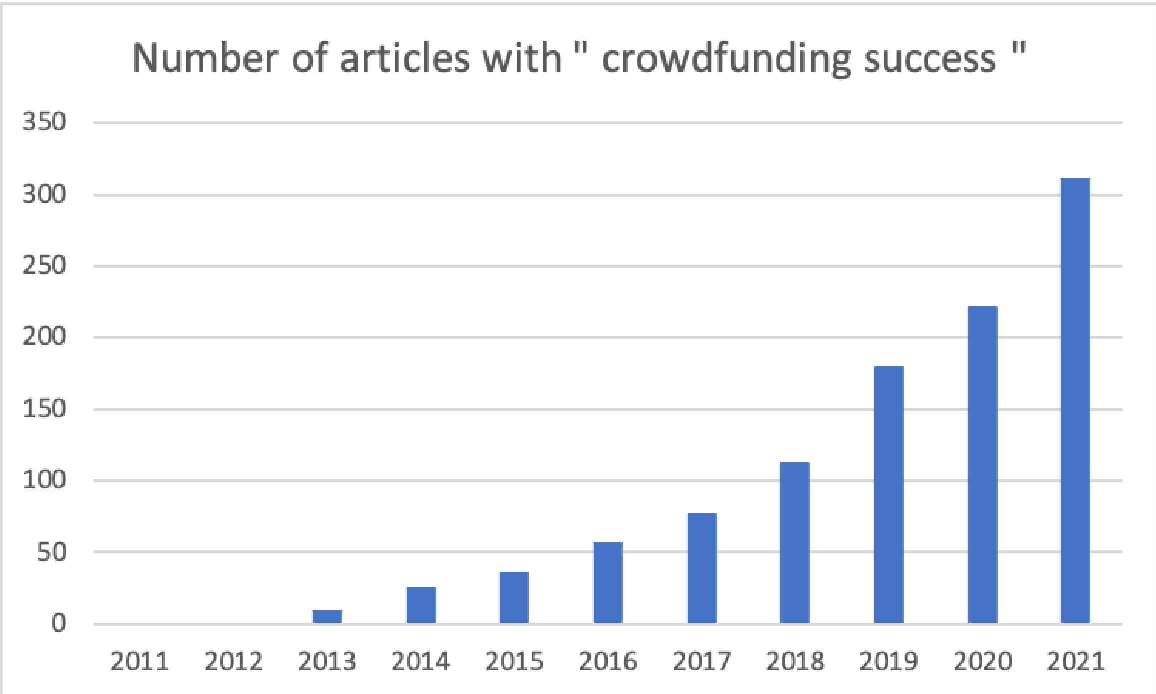



Fig. 3.2 Number of articles containing the term "crowdfunding success"

### 3.2.1 Equity Crowdfunding

The concept of crowdfunding was first introduced in the 1990s (Hoegen et al., 2018). It is the broader conception of crowdsourcing, which collects people's ideas for business activities (Belleflamme et al., 2014; Hervé and Schwienbacher, 2018). Much research has indicated the popularity and rapid growth of crowdfunding these years (Belleflamme et al., 2014) because of its distinct features compared to traditional investing ways. Likewise, also a number of researchers mentioned the importance of equity crowdfunding(ECF) in the development of new ventures, and there have been various ongoing academic discussions on equity crowdfunding in research area (Mochkabadi and Volkmann, 2020).

Pelizzon et al. (2016) review the historical development of equity crowdfunding, which briefly mentions its origins in the UK in 2011 and its subsequent growth in other countries, particularly in the context of startups and small and medium-sized enterprises (SMEs). The paper also analysed the differences between various business models of equity crowdfunding, illustrating the diversity of crowdfunding platforms and campaigns. According to their paper, equity crowdfunding has the ability to offer fresh prospects for both entrepreneurs and investors, as well as disrupt the predominance of conventional financing channels such as banks and venture capitalists. Kumar et al. (2016) and Son Turan (2015) described characteristics of equity crowdfunding. According to Kumar et al. (2016), equity crowdfunding is capable of overcoming conventional financing limitations. Additionally, it provides the potential for price discrimination, allowing issuers to set different prices for different groups of investors based on their willingness to pay and perceived value. Son Turan (2015) suggests that equity crowdfunding involves a range of stakeholders, including entrepreneurs, investors, crowdfunding platforms, regulatory bodies, and the wider community. The author highlights that each stakeholder faces specific risks and obstacles throughout the crowdfunding process. On top of that, Borello et al. (2015), Hornuf and Schwienbacher (2016) compared equity crowdfunding with conventional financing forms. Borello et al. (2015) studied 52 crowdlending and 67 equity crowdfunding platforms and found equity crowdfunding has some advantages over these traditional financing forms, such as lower transaction costs and greater flexibility, but also some disadvantages, such as higher risk and uncertainty. Whereas Hornuf and Schwienbacher (2016) mentioned that the relationship between equity crowdfunding investors and business angels is more likely to be a complement rather than a substitute. Similarly, some studies compared equity crowdfunding with the other types of crowdfunding (Belleflamme et al., 2014; Dorfleitner et al., 2017; Miglo, 2016). The potential of equity crowdfunding is also discussed (Dilger et al., 2017; Kim and De Moor, 2017; Stekli and Cali, 2020) and some even narrowed to regions and religion (Abdullah and Oseni, 2017; Mokhtarrudin et al., 2017; Nascimento and Querette, 2013). Part of the literature focuses on the impact of equity crowdfunding on entrepreneurs. For instance, a paper written by Walthoff-Borm et al. (2018) discusses the reasons behind the firm's decision to fundraise through equity crowdfunding. Moreover, Brown et al. (2019) mentioned the impact on the company through the equity crowdfunding process.

Table 3.1 Scholars about equity crowdfunding success

| Author(s) | Data Platform | Data Year | Data Truncation | Dependent Variable | model |
|---|---|---|---|---|---|
| Ahlers et al. (2015) | 104 from ASSOB (Australia) | 10/2006-10/2011 | NO | 1) Binary (Fully Funded=1, otherwise=0) 2) Number of investors 3) Funding amount 4) Speed of investment | OLS regression |
| Vulkan et al. (2016) | 635 from Seedrs(UK) | 07/2012-09/2015 | NO (from CTO) | Binary(Success=1, otherwise=0) | 1)Linear probability model 2)OLS 3)Quantile Regression |
| Block et al. (2018) | 71 from Seedmatch and Companisto (Germany) | 07/06/2012-27/04/2015 | NOT CLEAR | 1) Number of investment 2) Amount of capital pledged on a given day | 1)Negative binomial regression 2)OLS regression |
| Ralcheva and Roosenboom (2020) | 541 Crowdcube(UK) | 2012-03/2015 | NO | 1) Percentage of target amount raised 2) Number of investors | 1)Logit model 2)OLS regression 3)Negative binomial regression |
| Lukkarinen et al. (2016) | 60 from Invesdor (North Europe) | 05/2012-09/2014 | NO | 1) Number of investors 2) Amount raised | Multiple linear regression |
| Li (2016) | 49 from Dajiatou | Until 31/05/2015 | YES | 1) Ratio of fundraising completion 2) Fundraising speed 3) Number of followers | Independent-sample t-test |

| continue | | | | | |
|---|---|---|---|---|---|
| Vismara (2016) | 271 from Crowdcube and Seedrs | 02/2011 (Corwdcube) Since 2012(Seedrs) | NO | number of investors 2) funding amount (percentage of target amount collected) | 1)Negative binomial regression 2)OLS regression |
| Piva and Rossi-Lamastra (2018) | 129 from SiamoSoci (Itay) | 2012-01/02/2014 | NO | Dummy variable (=1 total pledges covered 100% of target capital, =0 if less) | Probit |
| Hornuf and Schmitt (2017) | 285 from (UK) | 01/08/2011-30/09/2016 | only observe successful ones | 1) dummy variable (=1 firm received additional funding, =0 otherwise) 2) dummy variable (=1 if firm insolvency liquidated dissolved, =0 otherwise) 3) time until follow-up funding 4) time until disappear | 1)Probit 2)Duration model 3)Binary outcome regression |
| Nitani and Riding (2017) | 141 (Germany) | 2012-01/02/2014 | NO | 1) Dummy variable (=1 amount reached, =0 otherwise) 2) Time until success | 1)Hierarchical logistic regression model 2)Proportional hazard models |

### 3.2.2 Equity Crowdfunding Success

Equity crowdfunding campaigns are either to succeed or to fail. Most equity crowdfunding platforms work on an all-or-nothing (AON) basis, which means the entrepreneur gets nothing if the goal is not achieved (Cumming et al., 2020). Therefore, a large proportion of equity crowdfunding literature is about the campaign strategies or factors that affect investors' behaviour and campaign success (Johan and Zhang, 2021). Table 3.1 shows most scholars worked on equity crowdfunding success factors, which is inspired by the idea of Mochkabadi and Volkmann (2020). Literature on campaign strategies and signals companies shown are organised into a few groups. For instance, a lot of research found that human capital such as top management team (TMT) size (Li, 2016; Giga, 2017), education, industry, and entrepreneurial experience (Nitani and Riding, 2017; Piva and Rossi-Lamastra, 2018) would influence the performance of equity crowdfunding campaigns. Gender of members in the management team has also been noticed by Cicchiello and Kazemikhasragh (2022) and Kleinert and Mochkabadi (2021), and literature mentioned gender bias and gap should not be ignored. Furthermore, Ahlers et al. (2015); Vismara (2016); Rossi et al. (2020) demonstrate that retaining equity (equity offered by the company to investors once the campaign succeeded) has a positive effect on the success of equity crowdfunding campaigns. Likewise, Ralcheva and Roosenboom (2020) predicted that a company that has retained equity, previous external financing and previous experience in accelerators is more likely to succeed in the equity crowdfunding campaign. It is also reported that more social capital, like a higher number of social network connections with entrepreneurs, leads to higher probabilities of success (Ahlers et al., 2015; Lukkarinen et al., 2016; Nitani and Riding, 2017; Vismara, 2016). Moreover, there is also literature about the characteristics of campaigns. Features like the number of investors and duration of campaigns affect the performance too (Lukkarinen et al., 2016; Vulkan et al., 2016). Recently, researchers found that the length of business descriptions has a positive effect on fundraising results (Dority et al., 2021; Johan and Zhang, 2021).

Concerning more specific results of some key literature, Cumming et al. (2019) mentioned that the venture shares type issued to investors has an impact on campaign success. They also found that the separation between ownership and control affects campaign success negatively, but this risk can be lower with experienced founders. As mentioned in the introduction, the UK has been the fastest-growing country for equity crowdfunding campaigns worldwide because of the clear regulatory framework that has been placed since 2011. Equity crowdfunding differs from the typical rewards-based crowdfunding with a much higher average amount pledged, average campaign goal, the existence of valuation of each of the projects and clear goal of the investors (backers) to gain a positive monetary return on their investment (Vulkan et al., 2016). Also, Ahlers et al. (2015) mentioned that small investors are often the target of start-up businesses on equity crowdfunding platforms because they do not normally have the ability to research and assess a potential investment extensively. Therefore clear signals to investors are found important. Ahlers et al. (2015)'s paper examines which crowdfunding project signals and attributes of venture quality are most likely to induce investors to commit financial resources in an equity crowdfunding context. The data of their paper shows the importance of the level of uncertainty to potential investors. For example, the amount of equity offered and whether the financial projections are provided. The percentage of board members with MBA degrees shows that human capital is important as well. However, intellectual capital (measured by patents) and social capital do not significantly impact funding success. For entrepreneurs, providing more information about risk can be interpreted as effective signals and increases the likelihood of funding success, and internal governance would increase the likelihood of attracting investors. Apart from that, Vismara (2016) mentioned that founders' behaviour is important to increase the probability of their ventures' success. The paper also showed the importance of entrepreneurs' social connections, which would help investors reduce information asymmetries and attract more investors and capital. Hornuf and Schwienbacher (2015) found that the dynamic pattern of crowd-investing is L-shaped, which means the funding rate increases rapidly at the beginning of a crowdfunding campaign and then levels off over time.[3] The authors extended their

[3]This results in an L-shaped curve when the amount of funding received is plotted against time.

studies on the dynamic effects of equity crowdfunding and mentioned that the L-shaped effect is under a first-come, first-served mechanism, and the U-shaped effect under a second-price auction (Hornuf and Schwienbacher, 2018).

Other than that, literature about research after the fundraising process extends the ECF area border. According to an interview-based study by Di Pietro et al. (2018), investors can contribute to companies' post-campaign performance, such as survival rates. A study of Hornuf et al. (2018) also mentioned that survival rates decrease with the successful subsequent equity crowdfunding campaign. Analysis of determinants of follow-up financing and firm failure after a successful campaign is applied by Hornuf et al. (2018), too. By comparing UK ventures with German ventures, they explore ventures of these two countries do have different possibilities of getting follow-up investments as well as different survival rates. Their study also shows the significant effect of ventures' characteristics on the follow-up findings. However, the results may be different between German ventures and UK ventures.

## 3.3 Descriptive Data Analysis

In this chapter, campaigns from two of the largest UK equity crowdfunding platforms are analysed. They are named Crowdcube and SEEDRS. The number of campaigns collected from these two companies is 416 and 478, respectively. Both data were collected by data collecting-tool with Python from as far as it could be obtained on the websites of the platforms until Oct 2019. As both platforms only show those successfully-funded campaigns,[4] which means the data is truncated here in our model due to the missing unsuccessful observations. Therefore we are applying the data to the truncated regression model in this chapter. As shown in Figure 3.3, the dependent variable ($log\frac{AmountRaised}{TargetAmount}$) is clearly truncated from below at zero since the distribution is uni-modal at zero, which demonstrates a straightforward truncated data and showing the mode of this distribution is very close to zero.

[4]We have contacted the platform, and they responded information on unsuccessful campaigns can not be provided. Email from the platform can be found in the Appendix.

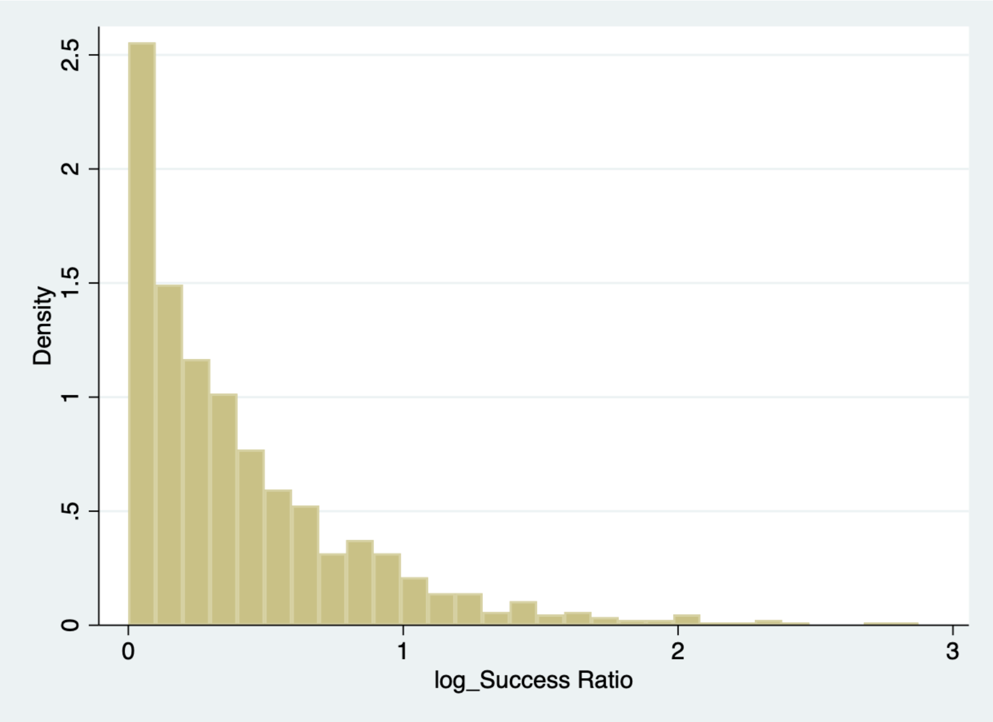



Fig. 3.3 Histogram of dependent variable $log(SuccessRatio)$

Summary statistics of all variables used in this chapter are presented in Table 3.2. It can be seen that the minimum success ratio is 1, which means that the raised amount exactly meets the target amount of the campaign.

| Variables | Observation | Mean | Std.Dev. | Min | Max |
|---|---|---|---|---|---|
| Target Amount(£) | 894 | 468074.5 | 1299508 | 688 | 2.20e+07 |
| Raised Amount(£) | 894 | 804794.5 | 2068988 | 813 | 3.93e+07 |
| Success Ratio | 894 | 1.750241 | 1.534204 | 1 | 23.94 |
| Equity Offered(%) | 894 | 11.46055 | 8.790206 | 0.05 | 94.6 |
| Valuation(£) | 894 | 9948458 | 7.20e+07 | 80 | 1.80e+09 |

Table 3.2 Summary Statistics of all Variables

To be noticed, there are also a few previous scholars who did research about these two platforms and unsuccessful campaign statistics appeared in their dataset as seen from Table 3.1. A paper written by Vulkan et al. (2016) analysed 636 campaigns of Seedrs in total, with both successful (33.9%) and unsuccessful ones. They got the data directly from the chief technology officer (CTO) of Seedrs, which means it is not directly available to investors.

More recently, a paper from Ralcheva and Roosenboom (2020) analysed both platforms, Crowdcube and Seedrs, with data on both successful and unsuccessful campaigns as well. However, their main dataset was from the professional data collecting company TAB, which provides data analytic services for the finance industry through AI and natural language processing. Moreover, the dataset of three UK equity crowdfunding platforms, Crowdcube, Seedrs and SyndicateRoom, appeared in a paper by Coakley et al. (2022), was also first obtained by TAB and then they matched it with data from Companies House, a government agency acting as the official registrar of UK firms, to obtain as much data as possible. Studies by Signori and Vismara (2016); Hornuf and Schmitt (2016) also get a part of their data from Companies House while analysing ECF success from the UK. Data from Companies House is publicly available but not highly consistent with the information shown on the equity crowdfunding platform itself. Apart from that, planned data was also applied to the models of papers by Vismara (2016) (Crowdcubes and Seedrs) and Nevin et al. (2017) (Crowdcube), which means they tracked/"subscribed" the live data from the platforms they were working on for a period of time therefore to get the statistics of both successful and unsuccessful data.

In this chapter, for having a direct first sight as investors, the data was collected directly from the equity crowdfunding websites like Hornuf et al. (2018) did in their paper of UK ECF platforms Crowdcube and Seedrs rather than third-party platforms. As only successfully funded campaigns are shown on these two specific websites and we did not track the live campaigns as the researchers mentioned above, data is then obviously truncated by the success threshold of the platforms. The data set collected includes the target amount of each campaign, the amount raised when the raising round ended, equity offered by the company to investors investing in their campaign, the number of investors and the pre-money valuation (the valuation before starting the fund-raising process) of the company. Two dependent variables are decided in this chapter to measure the success of equity crowdfunding campaigns like most research did. The first dependent variable is to see the investment sought by the company, which is the target the company wanted to reach. The other dependent variable is the ratio between the amount finally raised from investors and the target amount.

The other variables from data set are all independent variables used to see their effect on campaign success.

## 3.4 Methodology

In the previous literature on ECF, research methods vary between conceptual/theoretical, qualitative-like survey-based, to mainly quantitative analysis using regression. However, researchers have yet to use the truncated regression model. This is despite a number of studies clearly basing their analysis on truncated data sets. See Table 3.1. Therefore, in this chapter, we will start by explaining why the truncation in the data will cause least squares regression to give biased estimates (Maddala, 1983). We will go on to outline the truncated regression estimator, which is the appropriate estimator in this situation.

### 3.4.1 Truncation Bias

Hausman and Wise (1977) provide an explanation of why truncation in a data set gives rise to biased OLS estimates. In their example, the independent variable is years of education (of individuals) and the dependent variable is earnings, and anyone earning more than a certain level is missing from the data. Hence the data is upper truncated. Figure 1 of Hausman and Wise (1977) clearly illustrates how OLS estimates are biased in this situation, and in particular that the OLS slope is biased towards zero, that is, that the effect of education on earnings is underestimated.

Figure 3.4 shows a graph (based on simulated data) that is similar to Figure 1 of Hausman and Wise (1977). However, since here we are interested in lower truncation (firms with the lowest success rates are missing from the sample) we introduce lower truncation in Figure 3.4: all observations below the horizontal line are assumed to be missing from the sample. If all data were observed, OLS would give the “true line”, which is the dashed line in Figure 3.4. With the truncated data, OLS gives the higher dotted line. As with upper truncation, we see that the OLS slope is downward biased in a situation of lower truncation. The technical

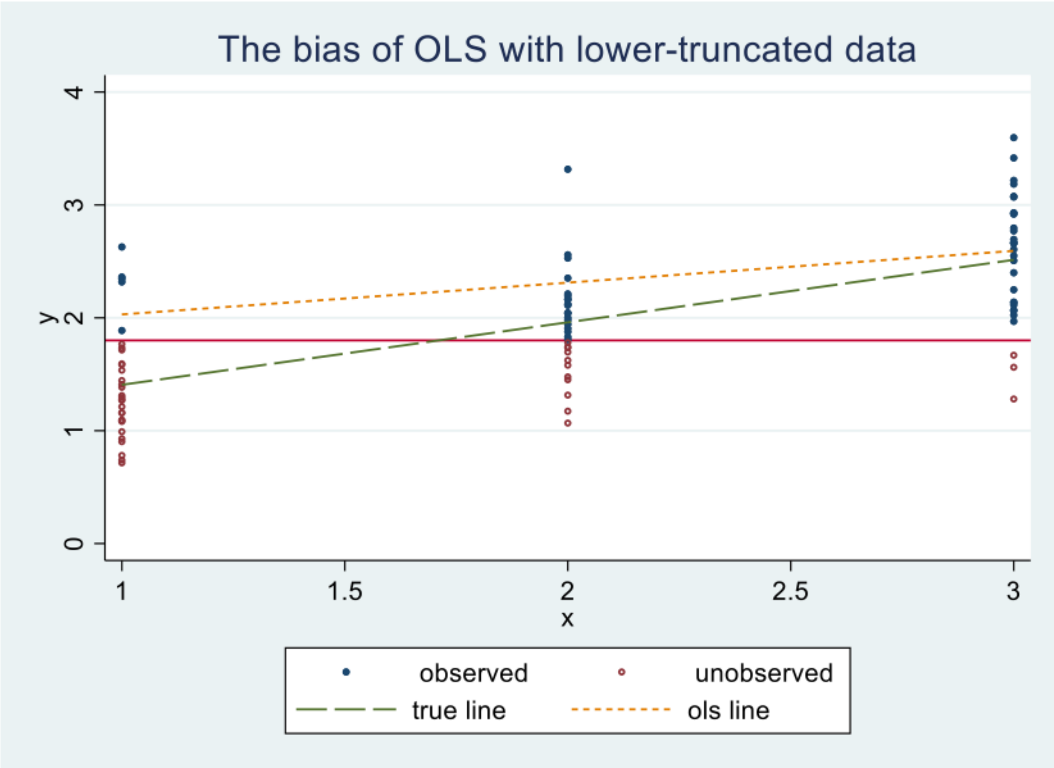



Fig. 3.4 The bias of OLS with lower-truncated data

reason for the bias is that the truncation gives rise to a strong negative correlation between the explanatory variable (x) and the equation error.

### 3.4.2 The Truncated Normal Distribution

One way of considering the truncation problem is to view the dependent variable $y_i$ as normally distributed with mean $x_i'\beta$ and variance $\sigma^2$, where $x_i'$ is a vector of explanatory variables, and $\beta$ as the corresponding vector of parameters. Such a density function is shown in Figure 3.5. Let us consider the maximum likelihood estimators of the parameters $\beta$ and $\sigma$ in this situation. To obtain these, we obtain the per-observation likelihood contribution which is based on the normal density function:

$$L_i(\beta,\sigma) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \tag{3.1}$$

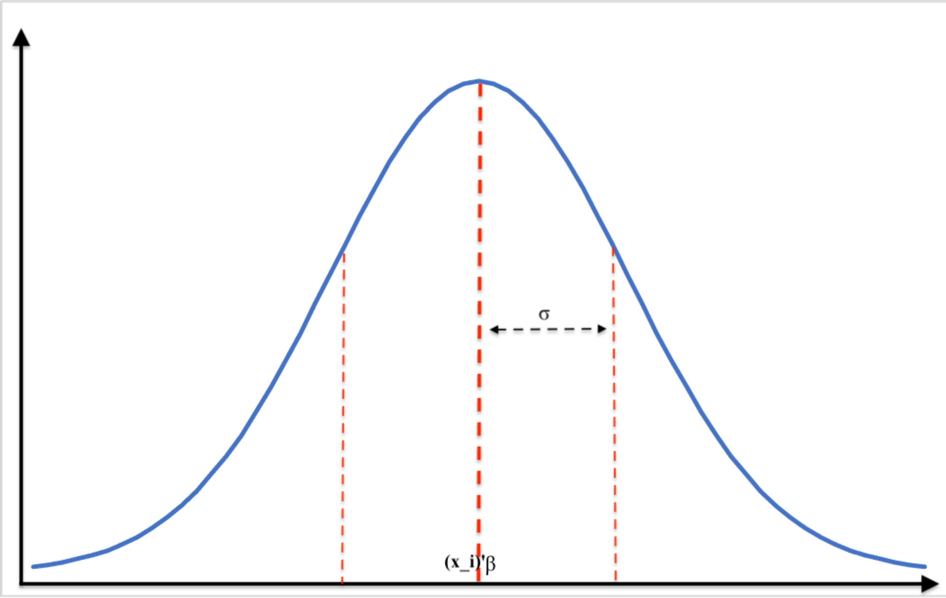



Fig. 3.5 Normal density function

where $\phi(.)$ is the standard normal density function. Then we would find the sample log-likelihood:

$$LogL(\beta,\sigma) = \sum_i ln(L_i(\beta,\sigma)) \tag{3.2}$$

Maximising eq.(3.2) with respect to $\beta$ and $\sigma$ gives maximum likelihood estimates (MLEs) of these parameters. And of course, these estimates are identical to the estimates obtained by applying linear (OLS) regression of $y_i$ on the variables contained in the vector $x_i$.

Now consider the situation in which the data set is lower truncated, so that we only observe data from the upper tail of the same distribution as considered above. How do we estimate the parameters in this situation? Here, the ML approach is required. However, the per-observation likelihood has a different structure to eq.(3.1).

As explained by Maddala (1983), if we used eq.(3.1), we would be basing the likelihood calculation on a function that is not a proper probability density function, because the area under the upper tail is clearly less than 1, while the area under a probability density function must be 1. Hence we need to scale the likelihood contributions by an appropriate amount.

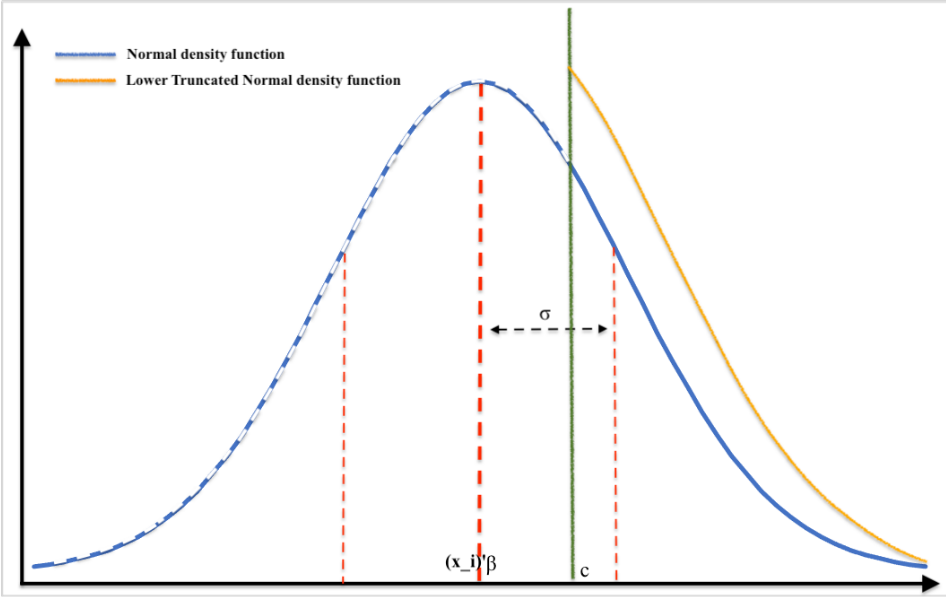



Fig. 3.6 Lower-Truncated Normal density function. The truncation point is $c$. Only values higher than $c$ are observed. Densities eq.(3.1) can be read off the blue curve. However, since the data is truncated, the densities eq.(3.3) need to be used instead, and these are read off the (higher) orange curve.

Specifically, if we divide each original density by the area of the tail, then we have a density function with the required property. See Figure 3.6, in which we show a normal distribution which is lower truncated at $c$, and we also show the truncated density which is necessarily higher than (i.e.shifted upwards) the non-truncated density, since both must have area 1. The per-observation likelihood contribution for the truncated regression model is therefore given by:

$$L_i(\beta, \sigma) = \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{c - x_i'\beta}{\sigma}\right)} \tag{3.3}$$

Note that the denominator of eq.(3.3) is the area of the upper tail of the normal distribution.

These concepts are used in later sub-sections where we consider the truncated regression model in the context of crowdfunding data.

### 3.4.3 Bias of the OLS Estimator under Truncation

To investigate the bias of the OLS estimator under truncation, consider the simplest possible case, in which there is only one independent variable ($x_i$), and both variables are in deviations from the mean. A similar case is considered in Section 6.9 of Maddala (1983).

$$\begin{aligned}
&y_i = \beta x_i + \varepsilon_i \\
&\varepsilon_i \sim N(0, \sigma^2) \\
&y_i \text{ observed if } y_i \geq c \\
&y_i \text{ unobserved if } y_i < c
\end{aligned} \tag{3.4}$$

Applying eq.(3.3) to eq.(3.4), taking logs, and summing over the sample, we obtain the sample log-likelihood:

$$LogL(\beta, \sigma) = -nln(\sigma) - nln(\sqrt{2\pi}) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \beta x_i}{\sigma}\right)^2 - \sum_{i=1}^{n} ln\left(1 - \Phi\left(\frac{c - \beta x_i}{\sigma}\right)\right) \tag{3.5}$$

Differentiating eq.(3.5) with respect to $\beta$ and setting to zero, we obtain one of the first-order conditions defining the MLE of $\beta$:

$$\frac{\partial LogL}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)x_i + \sum_{i=1}^{n}\frac{\phi\left(\frac{c-\beta x_i}{\sigma}\right)\left(\frac{x_i}{\sigma}\right)}{\left(1 - \Phi\left(\frac{c-\beta x_i}{\sigma}\right)\right)} = 0 \tag{3.6}$$

Rearranging eq.(3.6) we obtain:

$$\hat{\beta}_{MLE} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} + \frac{\sigma}{\sum_{i=1}^{n} x_i^2}\sum_{i=1}^{n}\frac{\phi\left(\frac{c-\beta x_i}{\sigma}\right) x_i}{\left(1 - \Phi\left(\frac{c-\beta x_i}{\sigma}\right)\right)} \tag{3.7}$$

Since the first term on the right-hand-side of eq.(3.7) is the OLS estimator of $\beta$, and the second term consists of only positive quantities (assuming x is always positive), we may re-write eq.(3.7) as:

$$\hat{\beta}_{MLE} = \hat{\beta}_{OLS} + \text{positive number}$$
$$\text{or}$$
$$\hat{\beta}_{OLS} = \hat{\beta}_{MLE} + \text{negative number} \tag{3.8}$$

Eq. (3.8) confirms that the bias of the OLS estimator of the slope is negative, in agreement with Figure 3.4 above. From (3.7), we see that this bias depends positively on $Var(\varepsilon_i)$ and negatively on $Var(x_i)$.

### 3.4.4 The Truncated Regression Model

Hausman and Wise (1977) introduce the truncated regression model which provides a means of obtaining (approximately) the "true line" even when the data is truncated.[5]

We assume that the relationship between the investment raised from the equity crowdfunding campaign of the company $i$ (the dependent variable), $y_i$, and the independent variables contained in the vector $x_i$ is of the form:

$$y_i = x_i{}'\beta + \varepsilon_i \tag{3.9}$$

$$\varepsilon_i \sim N(0, \sigma^2)$$
$$y_i \text{ observed if } y_i \geq y_i^{min}$$
$$y_i \text{ unobserved if } y_i < y_i^{min}$$

[5]This model can be estimated using the **truncreg** command in STATA.

where $i$ indexes campaigns, $y_i$ is log of amount raised for campaign $i$ and ${y_i}^{min}$ is log of investment sought for campaign $i$, i.e. ${y_i}^{min}$ is the log of the “target”. ${x_i}'$ is the vector of independent variables including equity offered ($e_i$), log of pre-money valuation of the company ($logv_i$) and a dummy variable $p_i$ showing whether the observation is from the platform crowdcube($p_i = 1$) or SEEDRS ($p_i = 2$).

The most important feature of the model described above is that the crowdfunding campaign is observed only if the amount raised meets the target, i.e. if $y_i \geq y_i^{min}$, otherwise it is absent from the data set.

Similarly to the previous sub-section, we construct the log-likelihood function by first considering the likelihood contribution for observation $i$:

$$L_i(\beta,\sigma) = \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - {x_i}'\beta}{\sigma}\right)}{1-\Phi\left(\frac{{y_i}^{min} - {x_i}'\beta}{\sigma}\right)} \tag{3.10}$$

The sample log-likelihood is then:

$$LogL(\beta,\sigma) = \sum_i ln(L_i(\beta,\sigma)) \tag{3.11}$$

$LogL$ in eq.(3.11) is maximised with respect to $\beta$ and $\sigma$ to obtain MLEs of these parameters.

As an alternative model, we consider the success ratio, defined as:

$$S_i = \frac{y_i}{{y_i}^{min}} \tag{3.12}$$

The dependent variable in this alternative model is actually:

$$s_i = ln\left(S_i\right) \tag{3.13}$$

$$s_i = {x_i}'\gamma + u_i$$

$$u_i \sim N(0, \eta^2)$$

$$s_i \text{ observed if } s_i \geq 0$$

$$s_i \text{ unobserved if } s_i < 0$$

One important difference between the two models is that in the first model (3.9) the lower truncation point is a variable (${y_i}^{min}$), while in the second model (3.13), the truncation point is a constant (0). Both cases are permitted within the **truncreg** command in STATA. The command has an option $ll(\cdot)$ for "lower limit". The argument of this option can be a variable or a single number.

### 3.4.5 A Test for Truncation Bias

An important question is how serious are the consequences of ignoring the truncation in the data and proceeding with estimation on the assumption that the data is not truncated. This question can be addressed using the Hausman test (Hausman, 1978). Aigner and Hausman (1980) apply the Hausman test to test for truncation bias. This approach is outlined here.

Let $\hat{\beta}_{ols}$ be the estimate of the vector $\beta$ obtained by applying the OLS estimator to eq.(3.9), and let $\hat{\beta}_{trunc}$ be that obtained by applying the truncated regression estimator defined in eq.(3.10). If we also obtain estimates of the variance matrices of the two estimates, $\hat{V}_{ols}$ and $\hat{V}_{trunc}$ respectively, then the Hausman test statistic is given by:

$$H = \left(\hat{\beta}_{trunc} - \hat{\beta}_{ols}\right)' \left(\hat{V}_{trunc} - \hat{V}_{ols}\right)^{-1} \left(\hat{\beta}_{trunc} - \hat{\beta}_{ols}\right) \tag{3.14}$$

and $H \sim \chi^2(k)$ under the null hypothesis of no truncation bias, where $k$ is the dimensionality of $\beta$. Hence, if the computed value of $H$ is greater than the critical value $\chi^2_{k,0.05}$, we may conclude that the two estimates are significantly different and that the estimate obtained using OLS is inconsistent as a result of truncation bias.

### 3.4.6 The Link Test

As a misspecification test, the link test will be used. The link test is a version of the well-known RESET test (Ramsey, 1969). The usefulness of the link test in microeconometric models has been established by Peters (2000).[6]

The link test is performed in two stages. The first stage is to estimate the model with the set of independent variables contained in the vector $x_i$ and generate the linear predictor:

$$\hat{p}_i = x_i{}'\hat{\beta} \tag{3.15}$$

The second stage is to estimate the model again, but using the two variables $\hat{p}_i$ and $\hat{p}_i{}^2$ instead of the variables contained in $x_i$. This model should also include an intercept. The link test statistic is the (asymptotic) t-test for testing the significance of $\hat{p}_i{}^2$ in this second model. If $\hat{p}_i{}^2$ shows significance, this indicates that the first model is misspecified in some way.

Peters (2000) has demonstrated that the test can be used as an "omnibus test". Hence, when it rejects, it simply indicates that there is some sort of misspecification. This could be in the form of a missing independent variable, or an incorrect distributional assumption, or a failure to account for a data feature such as truncation.

## 3.5 Results

Figure 3.7 and Figure 3.8 present the relationships between two dependent variables (log of Investment Amount and log of Success Ratio) and equity offered by applying The "Locally Weighted Regression" technique for smoothing data.[7] This technique was developed by Cleveland (1979), who recommended it is a very useful tool in exploratory data analysis for detecting non-linear relationships. The two figures strongly suggest that the relationships between equity offered and the two dependent variables are non-linear and non-monotonic.

[6] The test can be applied easily using the **linktest** command in STATA (StataCorp, 2021).
[7] It is available in STATA as "lowess".

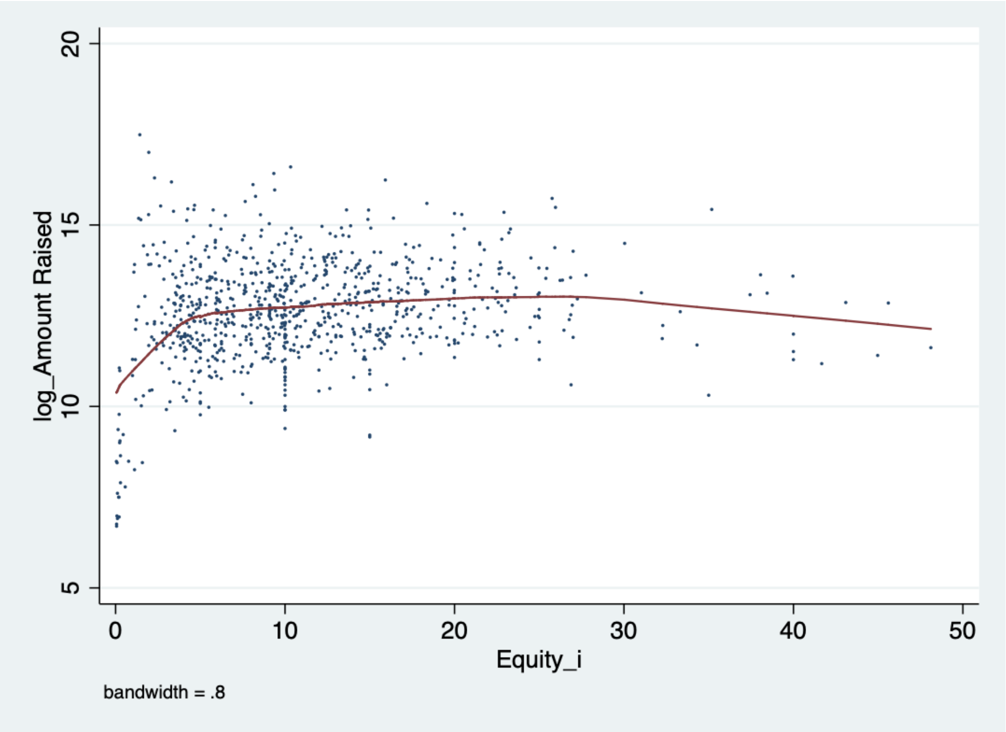



Fig. 3.7 Relationship between log of amount raised and equity offered

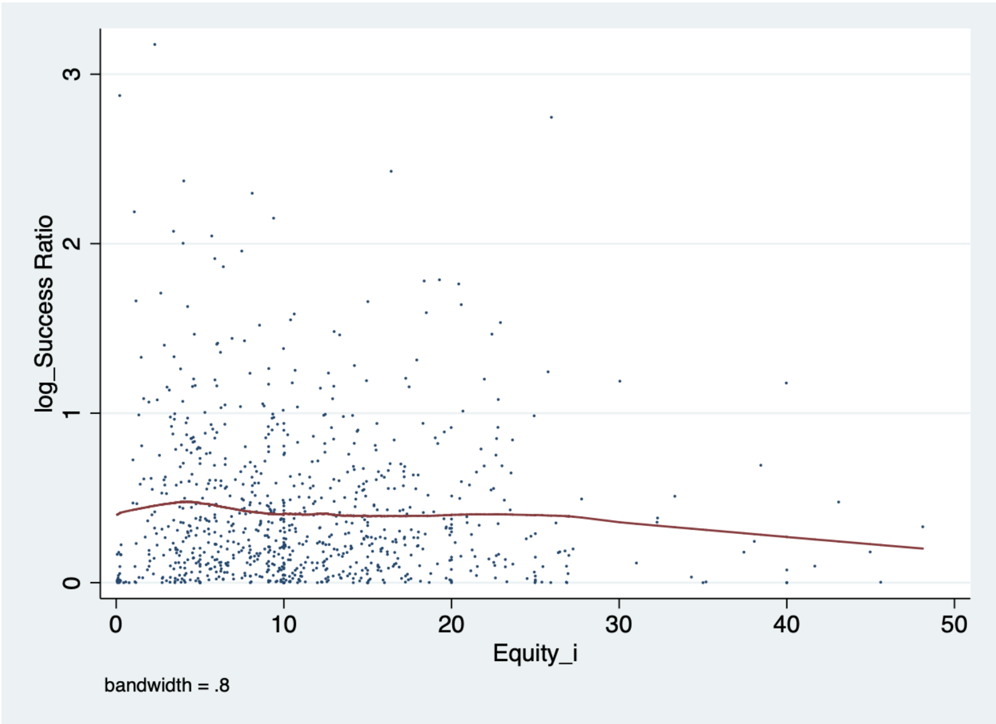



Fig. 3.8 Relationship between log of success ratio and equity offered

For this reason, the square and cube of the equity variable will be included in the models estimated.

Table 3.3 Results from OLS and TRM applied to model (3.9). Dependent variable: log of Amount Raised. In the truncated regression model (TRM), the (lower) truncation variable is the log of the target.

| VARIABLES | OLS | TRM |
|---|---|---|
| $equity$ | 0.213*** | 0.345*** |
| | (0.00897) | (0.0446) |
| $equity^2$ | -0.00403*** | -0.00961*** |
| | (0.000341) | (0.00201) |
| $equity^3$ | 2.37e-05*** | 6.08e-05*** |
| | (2.50e-06) | (1.45e-05) |
| $log\,(valuation)$ | 0.998*** | 1.498*** |
| | (0.0169) | (0.0932) |
| Seedrs(base: Crowdcube) | 0.423*** | -0.682** |
| | (0.0432) | (0.288) |
| Constant | -4.079*** | -14.05*** |
| | (0.274) | (1.713) |
| $\sigma$ | 0.624 | 0.987*** |
| | | (0.0696) |
| Observations | 894 | 894 |
| R-squared | 0.821 | |
| Hausman Test $(\chi^2(6))$(p-value) | 46.88(0.000) | |
| Link Test Statistic(p-value) | -0.00640(0.322) | 0.00520(0.662) |

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3.3 above shows both OLS regression results and truncated regression results of our first model (3.9), where the dependent variable is the investment achieved in each campaign.

Clearly, all variables have significant effects on the investment amount achieved, regardless of the truncated bias. However, the signs of the coefficients of equity crowdfunding platforms are different from the two regressions. It shows the truncation bias of our data in OLS. Both regressions show that the increase in the percentage of equity offered by companies would lead to a higher investment amount achieved, whereas the quadratic form of equity offered has a negative effect on investment achieved by the company. Graph 3.7 explains the reason. It shows that the relationship is non-monotonic, where the investment amount achieved by the company increases as the increase of equity offered until it reaches the highest point, and it then decreases when the equity offered continues increasing. The cubic form of equity provided has also been included because it is not monotonic, and the optimal percentage of equity offered is around 22%. Also, a higher pre-money valuation of the company, which means the valuation of the company before they start equity crowdfunding, would raise more funds because of this. The Hausman Test result between the two models, OLS and TRM, is also shown in the table. A strong significance of the chi-square number means a large and significant difference between the two models, emphasising the truncation bias and suggesting the application of the truncated regression model when the dataset is truncated.

Table 3.4 below shows the results of our alternative model (3.13) where the dependent variable is the log of the success rate. From Figure 3.8, it shows the change of direction of the slope, which indicates it is non-monotonic as well. So it sticks to model one and with an optimal figure of around 22% as well. According to other results in the table, there is a negative effect of Investment Sought on the success rate of the campaign, which means the higher the company's expectation is, it is less likely for them to successfully raise the fund. However, a higher pre-money valuation would increase the possibility for a company to raise the fund successfully. This indicates that if the valuation of the company before they go through the fundraising process on the platform is high, they would be more likely to raise enough amount of capital they would like. On top of that, data of campaigns collected from platform 2 SEEDRs performed worse than platform 1 Crowdcube. The reason for this might be different calculations of the information figures are applied in these two platforms. The coefficients between the two models here in Table 3.4 are obviously different. Also, from the

Hausman Test, it shows the chi-square number is strongly significant, which explains the noticeable difference between these two models and truncation bias in OLS.

Table 3.4 Results from OLS and TRM applied to model (3.13). Dependent variable: $ln(S_i)$ (log of Success Ratio). In TRM, the (lower) truncation point is zero.

| VARIABLES | OLS | TRM |
|---|---|---|
| *equity* | 0.0412*** | 0.408*** |
| | (0.00792) | (0.111) |
| $equity^2$ | -0.000848*** | -0.0114*** |
| | (0.000246) | (0.00369) |
| $equity^3$ | 5.92e-06*** | 7.24e-05*** |
| | (1.75e-06) | (2.47e-05) |
| $log(InvestmentSought)$ | -0.197*** | -1.166*** |
| | (0.0244) | (0.259) |
| $log(valuation)$ | 0.266*** | 1.719*** |
| | (0.0249) | (0.366) |
| Seedrs(base: Crowdcube) | -0.0647** | -0.720** |
| | (0.0323) | (0.337) |
| Constant | -1.404*** | -16.10*** |
| | (0.200) | (3.860) |
| $\sigma$ | 0.517 | 1.110*** |
| | | (0.0999) |
| Observations | 894 | 894 |
| R-squared | 0.419 | |
| Hausman Test $(\chi^2(7))$(p-value) | 52.98(0.000) | |
| Link Test Statistic(p-value) | 0.0252(0.679) | 0.0314(0.000) |

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In the analysis of this chapter, only a small number of explanatory variables have been used (equity ratio, target, company valuation, platform dummy). It may be seen as desirable to include more explanatory variables. With this in mind, I have extracted social media information on each company, and generated dummy variables for presence on the social media platforms: Facebook, Twitter, Instagram, and Linkedin. Unfortunately, these variables do not appear to have a significant effect on the level of success. The results including the social media variables are presented in Tables B.1 and B.2 of the Appendix.

### 3.5.1 Discussion of Link Test Results

The link test, described in Section 3.4.6, has been applied to all models. From Table 3.3, we see that both models, OLS and TRM are accepted on the evidence of the link test, when the dependent variable is the log of amount raised. From Table 3.4, we see that when the dependent variable is the log of the success ratio, OLS is accepted, but TRM is rejected. This suggests that there is some sort of misspecification when TRM is applied to the log of the success ratio. However, this is clearly a confusing result because the Hausman test tells us that there is strong evidence of truncation bias when OLS is used. Hence the two test results appear to be contradictory.

## 3.6 Conclusion

In this chapter, a fashionable way of investment, crowdfunding, has been introduced. Moreover, the main focus of this chapter is one of the four categories of crowdfunding called equity crowdfunding. Start-up companies sell their equity or bonds on an equity crowdfunding platform with details of their company, and investors could decide to buy and retain it like shares of the company. Previous literature did research on equity crowdfunding from different perspectives, and a number of recent ones also analysed the success factors of equity crowdfunding campaigns. As the data collection process is from an investor's point of view by scraping data directly from the equity crowdfunding websites rather than third party, data of unsuccessfully funded campaigns was not able to be accessed. Therefore truncated regression model is applied and compared with the OLS regression model in this chapter because of the truncation characteristic of our data. The results show that there is truncation bias in the OLS model, and according to the results obtained from the truncated regression model, the independent variables all have a significant impact on the success rate of raising enough funds for the campaign. Equity provided by the companies has a positive effect until it reaches an optimal number, and then the result would be negative, which means too much equity provided the investors might not have enough confidence in this campaign, so fewer people would take the risk to invest in this company. Investment sought, the target amount of a company also has a negative effect on the success fund-raising rate. The reason for this is possible because investors think the aim is too high to achieve, and they, especially non-experienced investors, would like to invest in those they think would be funded more easily. The company's pre-money valuation shows how it performed before they are on the platform. A higher value means they are more likely to be reliable and stable, so it attracts more investors. For the performance of two UK Equity Crowdfunding platforms Crowdcube and Seedrs, one of the reasons why Seedrs appear to perform worse is analysed to be distinct methods of calculation by the platform themselves and probably the location of scraping data from, since two websites show data in different layouts.

# Chapter 4

# Predicting the Success of Equity Crowdfunding Campaigns: Using Panel Data

## 4.1 Introduction

This chapter is an extension of the previous Chapter 3 by digging deeper into the data source from the secondary marketplace of one UK large equity platform, Seedrs. According to Seedrs (2021), their Secondary Market, which started in June 2017, provides an opportunity for early-stage private equity transactions. It operates as a bulletin board that enables investors to trade shares of businesses that have successfully raised investment on their platform under the Seedrs Nominee Structure during a Trading Cycle. The market opens for one week every month at a specific time, and investors are then able to make requests to other members to buy shares. Investors are benefited from this by investing in the business without it being under fundraising progress. Likewise, they also get an exit opportunity to sell the shares held anytime without holding them until the company is sold or going public. On the Seedrs website of the secondary market, information on companies having shares for investors to transact is shown. Each company/campaign consists of a sequence of "pitches", which shows the funding history of the company on Seedrs. Every pitch carries its own unique information, similar to a typical equity crowdfunding campaign. However, the data of the pitches within a company pertains to the performance of that particular company, and the information presented in each pitch may vary depending on the current situation of the respective company.

According to the literature mentioned in Chapter 3 Section 3.2, studies done by Di Pietro et al. (2018) and Hornuf and Schmitt (2017) investigated the post-fundraising process of equity crowdfunding. Mochkabadi and Volkmann (2020) emphasises the research gap of the ECF secondary market as there is only one German platform Innovestment, that has a second price auction similar to a secondary market before Seedrs. A paper by Lukkarinen and Schwienbacher (2020) found there is a positive effect on the number of investors and amount invested in the company when the company first entered the secondary market of equity crowdfunding, and the effect disappeared after the first 18 months of operating on the secondary market. These all indicate the importance of considering the follow-up pitch information of equity crowdfunding companies in the secondary market from the platform.

Although the data applied to the model of this chapter is from secondary market of Seedrs, this chapter still focuses the determinants of equity crowdfunding (ECF) success. However, different from Chapter 3, where data shows funded company section (only latest information of the campaign), and data in this chapter is from funded companies in secondary market of the platform, which shows data history of all successfully funded pitches of that campaign. So, it is panel truncated data collected for this chapter. Therefore, the truncated regression estimator in Chapter 3 is extended to the panel setting in this chapter.

Thus, the dependent variable is a continuous variable based on the amount raised and the target amount. Importantly, two features of the data are: (1) it is panel data because each campaign consists of a sequence of "pitches"; (2)it is truncated data because only successful pitches can be observed on the platform from the investors' point of view.

In summary, this chapter builds upon the analysis presented in Chapter 3 by delving deeper into the data source. We focus on Seedrs, a UK equity crowdfunding platform, where campaigns are comprised of a sequence of pitches. Therefore, in this chapter, we work with panel data since there are multiple observations per company. Our aim is to extend the truncated regression estimator to a panel setting, resulting in a new estimator that has not been previously discussed in literature. By applying this new estimator to the Seedrs dataset, we find that the results differ significantly from those obtained using random effects regression or truncated regression models alone. Our analysis reveals that crowdfunding success depends on several variables, including target amount, proportion of equity offered, and number of team members. Additionally, certain words used in pitch announcements have either a positive or negative impact on success. Through our analysis and use of the model estimates, we were able to determine optimal values for the continuous independent variables. These optimal values include a target amount of approximately £54,000, an offer of 23.60% equity, and a team size of 25 members.

The rest of the chapter is structured as follows: Section 4.2 describes the panel dataset. Section 4.3 shows the panel truncated regression model and several estimations applied, while Section 4.4 demonstrates the results and discussion of both linear and quadratic effects of models. Finally, Section 4.5 concludes.

## 4.2 Descriptive Data

As mentioned at the beginning, data analysed in this chapter is from the secondary market of Seedrs. Because only funded companies get the opportunity to go into the secondary market of the platform, the data of the company from Seedrs Secondary Marketplace is truncated. A histogram of our dependent variable (success ratio) in this Chapter is shown (Figure 4.1) to demonstrate the clear data truncation.

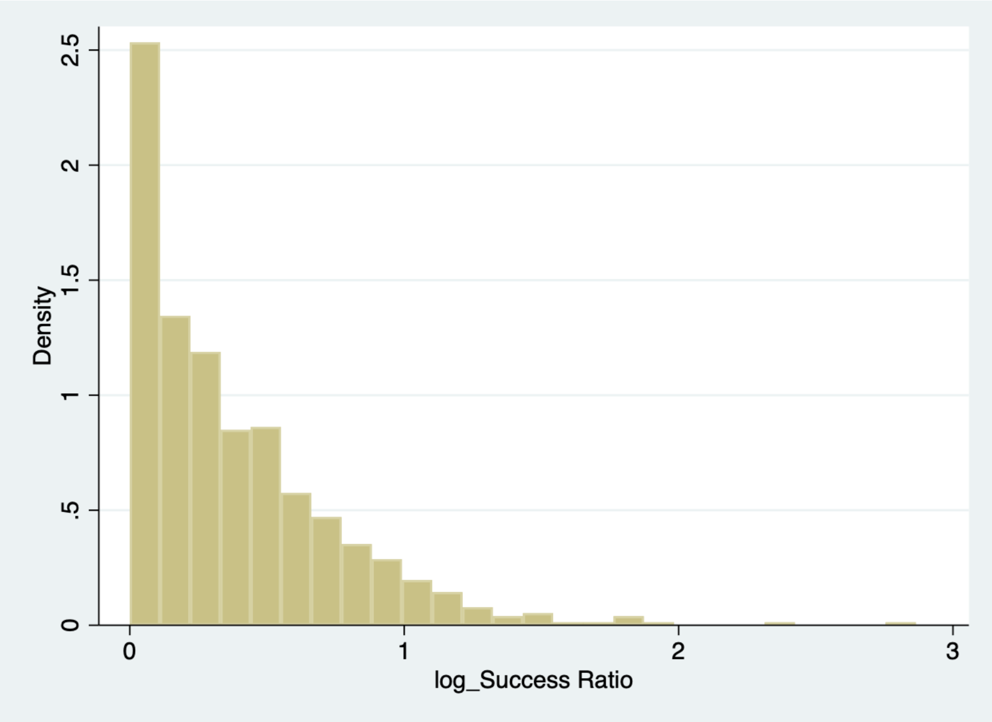



Fig. 4.1 Histogram of dependent variable $log\left(SuccessRatio\right)$

Data of 526 funded companies on the "secondary market" is obtained from Seedrs directly by using Python. Each company on the secondary market has at least one pitch, which means different rounds of their raising process (An example is shown in Figure 4.2 and Figure 4.3[1]).

Each pitch of the company shows the information about its target amount (amount of money the company would like to raise) and raised amount (amount of money the company finally got from investors) of that round, as well as the time that pitch/campaign has been approved by Seedrs Limited as a financial promotion and its closing date. Therefore, the duration between these two dates is also obtained as an independent variable to indicate the time length of raising progress for investors to consider before their investment. Summary statistics for all variables are presented in Table 4.1.[2]

[1] Figure 4.3 is the magnified depiction of the final segment of Figure 4.2.

[2] A few numbers of companies are raising in €, so we exchanged the amount to £ for consistency. However, not much influenced as the dependent variable is the log of the ratio between amounts.

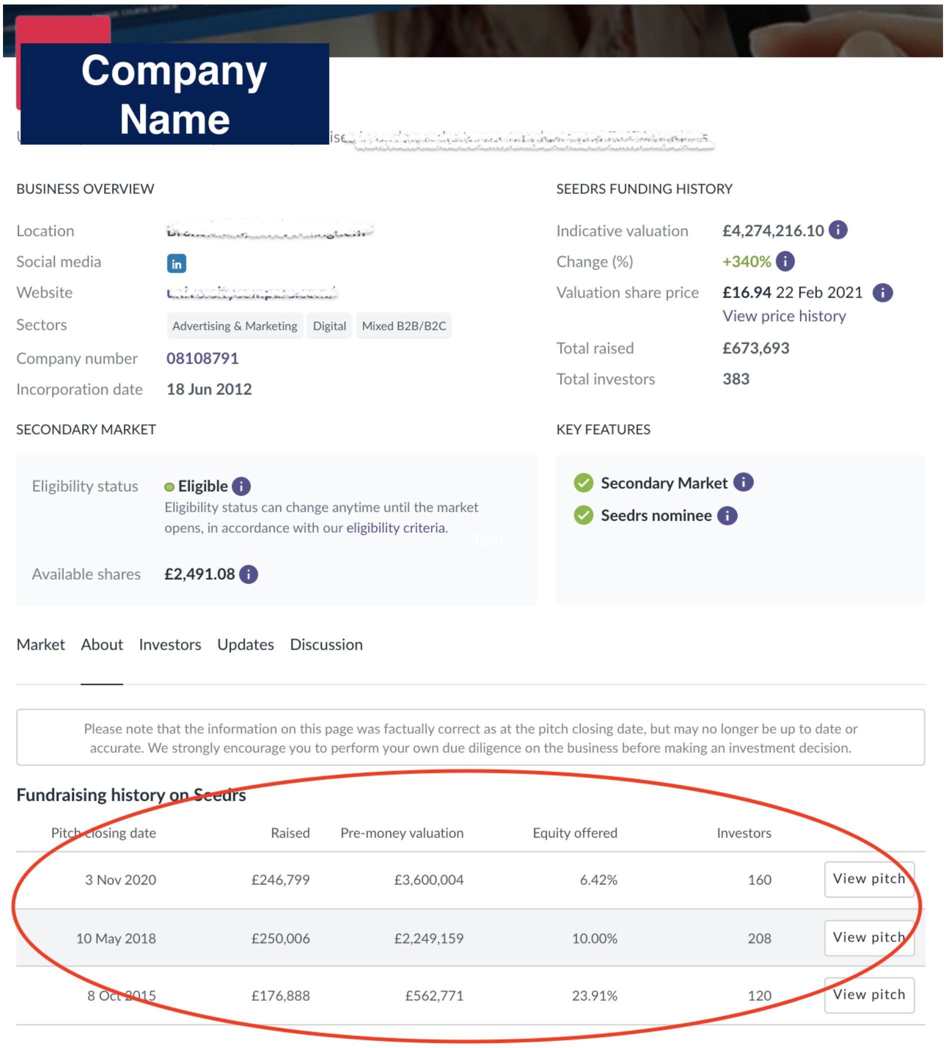
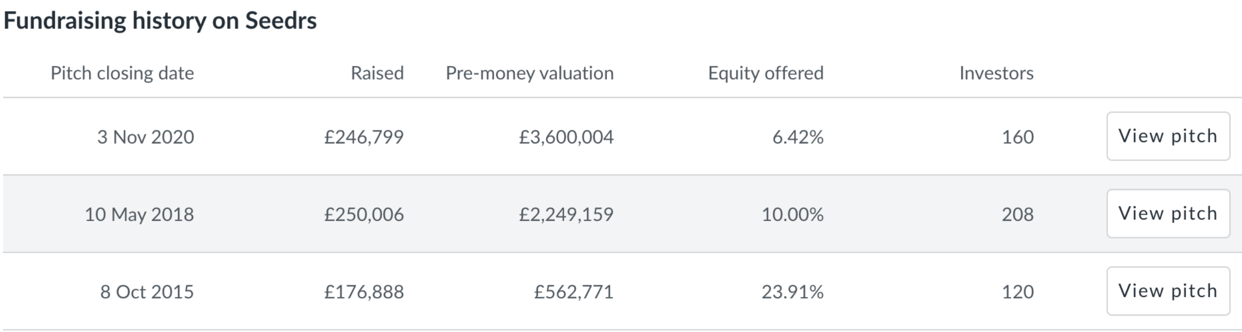



Fig. 4.2 Example of one company's page



Fundraising history on Seedrs

| Pitch closing date | Raised | Pre-money valuation | Equity offered | Investors | |
|---|---|---|---|---|---|
| 3 Nov 2020 | £246,799 | £3,600,004 | 6.42% | 160 | View pitch |
| 10 May 2018 | £250,006 | £2,249,159 | 10.00% | 208 | View pitch |
| 8 Oct 2015 | £176,888 | £562,771 | 23.91% | 120 | View pitch |

Fig. 4.3 Example of one company's pitches

| Variables | Observation | Mean | Std.Dev. | Min | Max |
|---|---|---|---|---|---|
| Target Amount(£) | 961 | 656165.6 | 2518949 | 688 | 7.12e+07 |
| Raised Amount(£) | 961 | 874670.7 | 2608690 | 813 | 7.12e+07 |
| Success Ratio | 961 | 2.420552 | 13.46842 | 1 | 386.4635 |
| Equity Offered(%) | 961 | 11.36638 | 8.724425 | 0.065 | 95.15 |
| Duration(day) | 961 | 22.84842 | 27.23163 | 1 | 170 |
| No. Team member | 961 | 5.902185 | 4.11081 | 1 | 26 |

Table 4.1 Summary of Variables

Apart from that, other independent variables are also included. According to the information and description provided by companies of their pitches, the number of team members in the company, which sector the company is in and certain words concluded from the descriptions are analysed to see if they have a significant effect on pitch success level. Words analysed in our model are mainly decided and selected by constructing the frequency of words from the descriptions of a number of selected companies by NVivo. An example of the word cloud made by the frequency of the words from the description is shown in Figure 4.4. Descriptive words like “sustainable” and “organic” are the main focus of analysis. Lastly, the dependent variable in this chapter is the logarithm of the ratio between the final raised amount and the target amount of pitch by the company, which is also applied in the previous Chapter 3. The relationship between the success ratio and each independent variable is shown in the graph 4.5 again obtained by applying the “Locally Weighted Regression” as Chapter 3 for smoothing data. As the relationship between each independent variable and success rate does not look monotonic, therefore the quadratic effect of independent variables on the dependent variable may appear here in our dataset.

Fig. 4.4 Word Cloud Example of one company

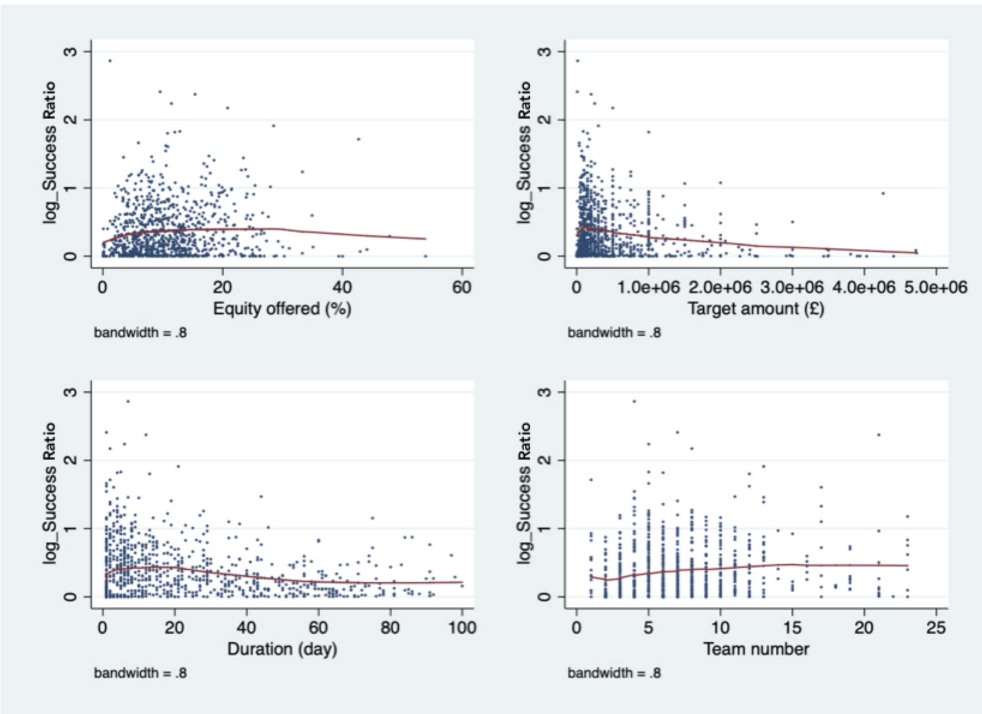




Fig. 4.5 Relationship between $log\,(SuccessRatio)$ and Independent Variables

## 4.3 The Panel Truncated Regression Model

In this section,[3] we develop a model that is suitable for modelling the truncated panel data set described in the last section. The model is a random-effects version of the truncated regression model. The truncated regression model is already well-established in the literature (Hausman and Wise, 1977).

### 4.3.1 Model and Notation

We assume that Company $i = 1, \ldots, n$ makes $t = 1, \ldots, T_i$ pitches. Let the amount raised by company $i$ in pitch $t$ be $y_{it}$, and let the target be $y_{it}^{min}$. The success of a pitch may be measured using the "success ratio" (independent variable) $S_{it} = y_{it}/y_{it}^{min}$. The dependent variable in our analysis will be $s_{it} = \ln S_{it}$. Note that $S_{it} \geq 1$ (or $s_{it} \geq 0$) indicates a successful pitch, while $S_{it} < 1$ (or $s_{it} < 0$) indicates an unsuccessful one.

We assume that the measure of success (the dependent variable) $s_{it}$ depends linearly on a set of $k$ independent variables contained in the $k \times 1$ vector $x_{it}$. The first element of $x_{it}$ is 1.

$$
\begin{aligned}
& s_{it} = x_{it}'\beta + u_i + \varepsilon_{it} \quad t = 1, ..., T_i \quad i = 1, ..., n \\
& u_i \sim N\left(0, \sigma_u^2\right) \\
& \varepsilon_{it} \sim N\left(0, \sigma_\varepsilon^2\right)
\end{aligned}
\tag{4.1}
$$

$\beta$ is a $k \times 1$ vector of parameters, the first of which is an intercept. $u_i$ is the company-specific effect, which captures unobserved differences between companies in their propensity to be successful in fundraising, and $\varepsilon_{it}$ is the equation error that is seen at the end of a regression equation.

[3] As stated in the Declaration at the start of the Thesis, this Section is co-authored with Prof. Peter Moffatt (UEA) and Dr. Simon Peters (Manchester).

### 4.3.2 Estimation

If $s_{it}$ were fully observed, we would proceed with the estimation of the model (4.1) using the standard random effects model, given that $u_i$ is not correlated with X. However, crucially, $s_{it}$ is only observed if $s_{it} \geq 0$; otherwise it is unobserved. For this reason, we require the random effects truncated regression model, estimated as follows.

Conditional on $u_i$, the likelihood contribution associated with an individual observation is:

$$L_{it} \mid u_i = \frac{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{s_{it}-x'_{it}\beta-u_i}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{x'_{it}\beta+u_i}{\sigma_\varepsilon}\right)} \tag{4.2}$$

The likelihood contribution for firm $i$ (still conditional on $u_i$) is:

$$L_i \mid u_i = \prod_{t=1}^{T_i}\left[\frac{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{s_{it}-x'_{it}\beta-u_i}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{x'_{it}\beta+u_i}{\sigma_\varepsilon}\right)}\right] \tag{4.3}$$

The marginal likelihood contribution for firm $i$ is then:

$$L_i = \int_{-\infty}^{\infty}\prod_{t=1}^{T_i}\left[\frac{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{s_{it}-x'_{it}\beta-u}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{x'_{it}\beta+u}{\sigma_\varepsilon}\right)}\right] f\left(u;\sigma_u\right)du \tag{4.4}$$

where $f\left(u;\sigma_u\right)$ is the normal density function for $u_i$, evaluated at $u$.

In the situation in which there is only one observation per firm, that is, $T_i = 1, \forall i$ and there are no $t$ subscripts, we may set $\sigma_u = 0$, and (4.4) simplifies to:

$$L_i = \frac{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{s_i-x'_i\beta}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{x'_i\beta}{\sigma_\varepsilon}\right)} \tag{4.5}$$

Equation (4.5) is the likelihood function for the cross-section truncated regression model developed by Hausman and Wise (1977). Thus we see that our model defined in (4.4) is an extension to the panel setting of this standard model already appearing in the literature and available in econometric software packages.[4]

For the panel version, the method of maximum simulated likelihood (MSL) is used, meaning that the integral in (4.4) is evaluated by finding the average over a sequence of suitably transformed Halton draws $(u_h, h = 1, \dots, H)$ (Train, 2009), as follows:

$$\hat{L}_i = \frac{1}{H} \sum_{h=1}^{H} \prod_{t=1}^{T_i} \left[ \frac{\frac{1}{\sigma_\varepsilon} \phi \left( \frac{s_{it} - x'_{it}\beta - u_h}{\sigma_\varepsilon} \right)}{\Phi \left( \frac{x'_{it}\beta + u_h}{\sigma_\varepsilon} \right)} \right] \tag{4.6}$$

Finally, the sample log-likelihood is obtained by taking the logarithm of (4.6), and summing over firms:

$$LogL = \sum_{i=1}^{n} ln \hat{L}_i \tag{4.7}$$

$LogL$ defined in (4.7) is maximised with respect to the parameters.[5]

### 4.3.3 Technical Issues

There are two important technical issues relating to the estimation of the panel truncated regression model outlined above. The first is that maximisation of the log-likelihood function is greatly facilitated by using the Olsen transformation (Olsen, 1978). Applied to the parameters of (4.4), this transformation amounts to: $\theta = 1/\sigma_\varepsilon$; $\eta = \beta/\sigma_\varepsilon$. The likelihood function expressed in terms of the transformed parameters is then:

$$L_i = \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} \left[ \frac{\theta \phi \left( \theta s_{it} - x'_{it}\eta - \theta u \right)}{\Phi \left( x'_{it}\eta + \theta u \right)} \right] f(u; \sigma_u)\, du \tag{4.8}$$

The sample log-likelihood is maximised with respect to the parameters $\theta$, $\eta$, and $\sigma_u$, and estimates of the structural parameters $\beta$, $\sigma_\varepsilon$, and $\sigma_u$ are recovered using the delta method (Oehlert, 1992).

[4] For example, the **truncreg** command in STATA estimates the model defined in (4.5). To our knowledge, no software package currently contains a routine for estimation of the panel version defined in (4.4).

[5] The log-likelihood function (4.7) is maximised using the ml routine in STATA. The STATA code is provided in the Appendix.

The second technical issue is that log-likelihood maximisation fails when only numerical derivatives are used in the maximisation algorithm. For this estimation problem to succeed, it is essential for analytical first derivatives of the log-likelihood function with respect to all parameters to be computed alongside the log-likelihood value itself.[6]

### 4.3.4 A Test for Truncation Bias

As in Chapter 3, we consider the question of how serious are the consequences of ignoring the truncation in the data and proceeding with estimation on the assumption that the data is not truncated. Once again, this question can be addressed using the Hausman test (Hausman, 1978). Aigner and Hausman (1980) apply the Hausman test to test for truncation bias in the cross-section setting. Here we extend the test to the panel setting.

Let $\hat{\beta}_{re}$ be the estimate of the vector $\beta$ obtained by applying the standard random effects estimator to (4.1), and let $\hat{\beta}_{ptrunc}$ be that obtained by applying the panel truncated regression estimator defined in (4.4). If we also obtain estimates of the variance matrices of the two estimates, $\hat{V}_{re}$ and $\hat{V}_{ptrunc}$ respectively, then the Hausman test statistic is given by:

$$H = \left(\hat{\beta}_{ptrunc} - \hat{\beta}_{re}\right)' \left(\hat{V}_{ptrunc} - \hat{V}_{re}\right)^{-1} \left(\hat{\beta}_{ptrunc} - \hat{\beta}_{re}\right) \tag{4.9}$$

and $H \sim \chi^2(k)$ under the null hypothesis of no truncation bias, where $k$ is the dimensionality of $\beta$. Hence, if the computed value of $H$ is greater than the critical value $\chi^2_{k,0.05}$, we may conclude that the two estimates are significantly different and that the estimate obtained from standard random effects is inconsistent as a result of truncation bias.

Since it is not straightforward to obtain the estimated variance matrix $\hat{V}_{ptrunc}$, a simpler version of the Hausman test is also valid. This version compares the ols regression estimator ($\hat{\beta}_{ols}$) and the (cross-section) truncated regression estimator ($\hat{\beta}_{trunc}$) , but using cluster-robust variance matrices for both ($\hat{V}_{ols,c}$ and $\hat{V}_{trunc,c}$ respectively). This version of the test is therefore

[6] To build analytic first derivatives into the STATA ml routine, method d1 is used in place of method d0.

given by:

$$H^{\star} = \left(\hat{\beta}_{trunc} - \hat{\beta}_{ols}\right)' \left(\hat{V}_{trunc,c} - \hat{V}_{ols,c}\right)^{-1} \left(\hat{\beta}_{trunc} - \hat{\beta}_{ols}\right) \tag{4.10}$$

### 4.3.5 Post-estimation

Having estimated the panel truncated regression model defined in (4.4) it is possible to obtain posterior estimates of the random effect term for each company. The posterior estimate for company $i$ is given by:

$$\hat{u}_i = \frac{\int_{-\infty}^{\infty} u \prod_{t=1}^{T_i} \left[ \frac{\frac{1}{\sigma_{\varepsilon}} \phi\left(\frac{s_{it} - x_{it}'\beta - u}{\sigma_{\varepsilon}}\right)}{\Phi\left(\frac{x_{it}'\beta + u}{\sigma_{\varepsilon}}\right)} \right] f\left(u; \sigma_u\right) du}{L_i} \tag{4.11}$$

where $L_i$ is the likelihood contribution for company $i$, defined in (4.4).

Note that (4.11) gives a *ceteris paribus* ranking of companies by successfulness. That is, the ranking is obtained in a way that controls for all of the pitch-characteristics that are included as independent variables in the model.

## 4.4 Empirical Analysis

This section presents the results of models from different aspects. Firstly, we will discuss the linear effect of models on ECF pitch success. Then followed by a textual analysis part as several words are included in the model. In the end, we estimate the optimal values of variables and show a list of company names according to their performance.

### 4.4.1 Linear Specification

As we can see from Table 4.2, where four regression results are shown. They are Random Effects Regression with Panel Data, Truncated Regression with Clustering on Cross-Sectional Data and finally our Truncated Regression Model on Panel Dataset without and with selected words to analyse. Comparing with all of them, the coefficients of each independent variable

Table 4.2 Results from various models: linear effects only. Dependent variable: $s_{it}$ (log of success ratio). In truncated models, the lower truncation point is zero.

| VARIABLES | Random Effects Panel Model | Cross-Section Truncated Regression with Clustering | Panel Truncated Regression | Panel Truncated Regression |
|---|---|---|---|---|
| $log(target)$ | -0.0728*** | -0.451*** | -0.297*** | -0.232*** |
| | (0.0131) | (0.0848) | (0.0640) | (0.0485) |
| $log(duration)$ | -0.104*** | -0.614*** | -0.347*** | -0.307*** |
| | (0.0102) | (0.622) | (0.0607) | (0.0451) |
| $log(equity)$ | 0.0979*** | 0.598*** | 0.451*** | 0.299*** |
| | (0.0187) | (0.118) | (0.0947) | (0.0669) |
| team number | 0.0179*** | 0.100*** | 0.0645*** | 0.0503*** |
| | (0.00322) | (0.0206) | (0.0154) | (0.0116) |
| Data & Analytics Sector | -0.168* | -1.410* | -0.979** | -0.676* |
| | (0.0873) | (0.736) | (0.477) | (0.371) |
| Energy Sector | 0.129* | 0.939** | 0.399 | 0.463* |
| | (0.0745) | (0.467) | (0.282) | (0.236) |
| Finance & Payments Sector | 0.0753** | 0.451** | 0.225* | 0.248** |
| | (0.0366) | (0.228) | (0.135) | (0.116) |
| “health/healthy” | 0.104*** | 0.487*** | - | 0.256*** |
| | (0.0320) | (0.185) | | (0.0961) |
| “organic” | 0.544 | 0.428** | - | 0.222** |
| | (0.0373) | (0.219) | | (0.108) |
| “planet” | 0.149** | 0.660* | - | 0.326* |
| | (0.0690) | (0.354) | | (0.178) |
| “quality” | 0.0617** | 0.365** | - | 0.196** |
| | (0.0277) | (0.169) | | (0.0848) |
| “data” | 0.0848*** | 0.505*** | - | 0.252*** |
| | (0.0293) | (0.173) | | (0.0883) |
| “entertainment” | -0.108 | -1.181** | - | -0.562* |
| | (0.0679) | (0.568) | | (0.287) |
| “information” | -0.885*** | -0.693*** | - | -0.336*** |
| | (0.0306) | (0.209) | | (0.107) |
| Constant | 1.188*** | 4.503*** | 2.765*** | 2.320*** |
| | (0.153) | (1.01) | (0.621) | (0.508) |
| $\sigma_u$ | 0.333 | - | 0.444*** | 0.409*** |
| | | | (0.0760) | (0.0597) |
| $\sigma_\varepsilon$ | 0.337 | 0.824*** | 0.572*** | 0.493*** |
| | | (0.149) | (0 .0667) | (0.0491) |
| Log Likelihood | | 62.01 | 57.12 | 79.52 |
| $AIC = 2k - 2LogL$ | | -92.02 | -94.24 | -125.04 |
| Link Test Statistic (p-value) | 4.00(0.00) | 1.05(0.29) | 2.33(0.02) | 2.82(0.01) |
| n: Number of Companies | 526 | 526 | 526 | 526 |
| T: Number of Pitches/Company | 1.83 | 1.83 | 1.83 | 1.83 |

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

changed when we applied the data into the truncated regression model for panel data and the results are significant as well. Table 4.2 also shows that normal truncated regression worked because of the significance of the second regression model (Cross-Sectional Truncated Regression), compared with the random effect regression results, which means taking into account the truncated nature of the dependent variable impacts on the estimates. The results demonstrate target amount, equity offered by the company and number of team members all have strong positive effects on the success rate of each fundraising pitch for companies, whereas the duration of the pitch has a negative effect on the success rate. This means when investors are trying to invest in this pitch, if they find out that the target amount, equity offered by the company and team numbers of that pitch are high on the pitch information website, then the pitch is more likely to succeed if the non-monotonic effect is not considered. Likewise for explaining the negative effect of the duration between the approved date and the pitch closing date. Therefore, when investors are willing to invest in any pitch, they could also take the date into consideration. If the date they are investing is too far from the approved date of that pitch, that would be relatively less likely for them to succeed compared with the newly approved pitches. Moreover, among all sectors from the secondary market of Seedrs, results found that the Data & Analytics sector, Energy Sector and Finance & Payment Sector have an effect on camping success when regarding the ratio between the raised amount and target amount as the success rate (dependent variable). However, companies in the Data & Analytic sector are likely to have a higher possibility to fail compared to companies in other sectors. The coefficients are much higher than the random effect for the cross-sectional truncated regression model, which again shows the truncation bias and it is a bit lower for the truncated regression model for panel data. This means the results would be more robust if we consider the specific type of dataset applied to the model.

In Section 4.3.4 above we outlined the use of the Hausman test to test for truncation bias in the panel data context. This test compares two estimators, linear random effects, and panel truncated regression. We also outlined a simpler version of the test that compares OLS and cross-section truncated regression, but using cluster-robust variance estimates (hence allowing for the panel structure). When we apply the simpler version of the test to this data set, we obtain a test statistic $(\chi^2(16))$ of 18.307, and a p-value of 0.0003. This indicates that there is strong evidence of truncation bias in any model that fails to account for truncation.

### 4.4.2 Textual Analysis

The final column of Table 4.2 presents results from the panel truncated regression model with a number of "word dummies" included. These are 0/1 dummy variables, indicating whether a particular word appears in the pitch. As mentioned in Section 4.2, a word cloud of all pitches was first used to identify a set of commonly appearing words. These words were included in the model, and only those with a significant effect were retained. Seven word dummies are included in the model. The importance of these words is confirmed using a likelihood ratio (LR) test. The LR statistic is computed as $LR = 2 \times (79.52 - 57.12) = 44.8$, which is considerably greater than the 1% critical value with 7 degrees of freedom, $\chi^2_{7,0.01} = 14.07$. This confirms that the word dummies have strongly significant joint effects. Also, the AIC for the model with word dummies is the lowest of all the models, confirming that adding word dummies improves the fit of the model, even allowing for an increase in the number of parameters.

Words like "health/healthy", "organic", "quality", which relate to people's lifestyles, all appear to have a positive effect on pitch success. Words like "planet", which relates to the environment, also appear to have a positive effect. These all suggest that investors attach great importance to the development of healthy lifestyles and the environmental sustainability of the planet. The word "data" has a positive effect, and this suggests that investors have more confidence in the ECF pitch if the company provides information about data management or if the product is closely related to data. By contrast, the word "information" appears to have a negative effect. The implication is that a company should use "data" in preference to "information" in any situation in which the meaning is the same. Another word that appears to have a negative effect is "entertainment".

These results are potentially very useful to companies in the process of designing a campaign pitch. However, the implication is not that companies can increase success by simply including certain words in the pitch. The implication is that investors are attracted by products with particular features, and the words in the pitch provide the necessary information to investors on whether these features are present. Hence the words are acting as proxies for the product type. Notice that company sector dummies are also present in the model. The word dummies have stronger significance than the company sector dummies.

### 4.4.3 Optimal Value Estimation

In Table 4.3, the first column reproduces the final column of Table 4.2. The second column is the same model, but with quadratic terms in all continuous variables ($log\,(target)$, $log\,(equity)$, $log\,(duration)$ and $TeamNumber$). The importance of the quadratic terms is confirmed using a likelihood ratio (LR) test. The LR statistic is computed as $LR = 2 \times (96.11 - 79.52) = 33.18$, which is considerably greater than the 1% critical value with 4 degrees of freedom, $\chi^2_{4,0.01} = 13.28$. This confirms that the quadratic terms have a strongly significant joint effect. According to the AIC, this model, with quadratic terms and word dummies all included, is the best-fitting model of all models estimated on this data set.

For all of the continuous variables, the sign of the quadratic term is negative, indicating an inverted u-shaped effect, although in the case of team number, the quadratic term is insignificant.

It is possible to derive the optimal value of each independent variable by dividing the coefficient of the linear term by two times the coefficient of the quadratic term, and reversing the sign. We can then obtain a confidence interval for the optimal value using the delta method (Oehlert, 1992).[7] These optimal values are presented in Table 4.4. Noticeable differences are seen in the optimal values when they are obtained from different models. For example, based on the random effects model (ignoring truncation) the optimal equity ratio is 63%, while for the panel truncated model, it goes down to 23.6%. Another difference, seen from the width of the confidence intervals, is that the truncated models give much more accurate estimates of the optimal values than the model that ignores truncation.

[7] The delta method can be applied easily using the **nlcom** command in STATA.

Table 4.3 Results from two models: with and without quadratic effects. Dependent variable: $s_{it}$ (log of success ratio)

| VARIABLES | Panel Truncated Regression | Panel Truncated Regression |
|---|---|---|
| $log\,(target)$ | -0.232*** | 1.843** |
| | (0.0485) | (0.0813) |
| $log\,(target)^2$ | - | -0.0845** |
| | | (0.0330) |
| $log\,(duration)$ | -0.307*** | 0.0874 |
| | (0.0451) | (0.114) |
| $log\,(duration)^2$ | - | -0.0931*** |
| | | (0.0272) |
| $log\,(equity)$ | 0.299*** | 0.778*** |
| | (0.0669) | (0.282) |
| $log\,(equity)^2$ | - | -0.123** |
| | | (0.0621) |
| $TeamNumber$ | 0.0503*** | 0.0781*** |
| | (0.0116) | (0.0295) |
| $TeamNumber^2$ | - | -0.00155 |
| | | (0.00143) |
| Data & Analytics Sector | -0.676* | 0.520 |
| | (0.371) | (0.372) |
| Energy Sector | 0.463* | 0.538*** |
| | (0.236) | (0.241) |
| Finance & Payments Sector | 0.248** | 0.203* |
| | (0.116) | (0.113) |
| "health/healthy" | 0.256*** | 0.249*** |
| | (0.0961) | (0.0959) |
| "organic" | 0.222** | 0.197* |
| | (0.108) | (0.108) |
| "planet" | 0.326* | 0.286 |
| | (0.178) | (0.175) |
| "quality" | 0.196** | 0.198** |
| | (0.0848) | (0.0857) |
| "data" | 0.252*** | 0.190*** |
| | (0.0883) | (0.0878) |
| "entertainment" | -0.562* | -0.546* |
| | (0.287) | (0.291) |
| "information" | -0.336*** | -0.226*** |
| | (0.107) | (0.109) |
| Constant | 2.320*** | -11.085** |
| | (0.508) | (5.015) |
| $\sigma_u$ | 0.409*** | 0.340*** |
| | (0.0597) | (0.0699) |
| $\sigma_\varepsilon$ | 0.493*** | 0.507*** |
| | (0.0491) | (0.0528) |
| Log Likelihood | 79.52 | 96.11 |
| $AIC = 2k - 2LogL$ | -125.04 | -150.22 |
| Link Test Statistic (p-value) | 2.82(0.01) | 1.89(0.06) |
| n: Number of Companies | 526 | 526 |
| T: Number of Pitches/Company | 1.83 | 1.83 |

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

| VARIABLES | Random Effects Panel Model | Cross-Section Truncated Regression with Clustering | Panel Truncated Regression |
|---|---|---|---|
| Equity Offered (%) | 63.38<br>[-161.949 , 288.699] | 25.25<br>[-0.889 , 51.395] | 23.60<br>[-2.302 , 49.507] |
| Target Amount (£) | 46136.63<br>[-20731.09 , 113004.4] | 9544.59<br>[-32204.36 , 51293.54] | 54275.59<br>[-13199.57 , 121750.8] |
| Duration (day) | 0.731<br>[-0.774 , 2.235] | 1.659<br>[0.259 , 3.059] | 1.599<br>[0.00333 , 3.195] |
| No. Team member | 34<br>[-27.066 , 94.885] | 41<br>[-66.868 , 148.271] | 25<br>[-4.684 , 55.103] |

Table 4.4 Point Estimates of Optimal Values (interval estimates in parentheses; obtained using delta method.)

The optimal target amount appears to be around £54,000. The optimal value for the duration between the approved date and the pitch closing date is 1.6 days which is dramatically lower than expected. A possible explanation for this is that companies are able to start raising capital for their later pitches before they get approved.[8] Lastly, although the quadratic term for the number of team members in the company is not significant, it is still useful to know that the optimal number of team members is estimated to be 25.

There are theoretical explanations for some of these quadratic effects. In the case of equity ratio, Vismara (2016) suggests that entrepreneurs' willingness to invest in their own project signals quality, and this signal is important because the entrepreneur has more information than the investor on the quality of the project. If investors pay attention to this signal, then the equity ratio (proportion of equity offered) is predicted to have a negative effect on the level of success. On the other hand, if the proportion of equity offered is very low, this might be perceived as a signal that the external investors are not regarded as important to the success of the project, and this might put investors off. Investors want to feel that they are making an important contribution. The results support both of these hypotheses because there is strong evidence of an inverted-u shaped effect.

[8] However, this is just a guess, and more research is required to explain this surprising result.

In the case of the number of Team members Hornuf and Schmitt (2017) have hypothesized as follows: “On the one hand, starting a business alone can be difficult and cumbersome because of lack of competences and capacity constraints. On the other hand, the larger the management team of the start-up becomes, the more likely are disputes among management team members to arise”. Since our coefficient of the number of team members is positive and significant, our results support the first part of this hypothesis. However, our results do not support the second part of the hypothesis because the coefficient of the quadratic term is not significant.

### 4.4.4 Discussion of Link Test Results

The link test, described in Section 3.4.6 of Chapter 3, has again been applied to all models.

From Table 4.2, we see that all models appear to be misspecified except the cross-section truncated regression model. This result is confusing because this model does not allow for between-company heterogeneity, and we therefore expect to reject it. From Table 4.3, we see that the final model, that is the model including quadratic terms and word dummies, passes the link test, with a p-value of 0.06. This is a welcome result, because it means that the model considered (a priori) to be the best model in the thesis also appears to fit the data.

### 4.4.5 Posterior Random Effects

Table 4.5 shows the posterior random effects, obtained using the formula (4.11), for some of the companies. The companies that are shown are the ones with the eleven highest posterior random effects (i.e. the eleven most successful companies), and the ones with the ten lowest posterior random effects (i.e. the ten least successful).

Some types of companies (e.g. Finance & Payments, SaaS/PaaS) appear both at the top and at the bottom of the ranking, so it is hard to discern meaningful patterns from this table.

| | Company Names | Sector | $\hat{u}_i$ |
|---|---|---|---|
| Most Successful | Ziglu | Finance & Payment | 0.999 |
| ↑ | YellowDog | SaaS/PaaS | 0.972 |
| | MacRebur | Automotive & Transport | 0.950 |
| | Rubies in the Rubble | Food & Beverage | 0.862 |
| | cycle.land | Automotive & Transport | 0.781 |
| | PURE SOUTH PRESS | Food & Beverage | 0.756 |
| | Ticketpass | SaaS/PaaS | 0.746 |
| | DMALINK | Finance & Payment | 0.713 |
| | Holly and Beau | Clothing & Accessories | 0.710 |
| | cloudplan | SaaS/PaaS | 0.709 |
| | Nurturey | Healthcare | 0.692 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| | Audiopi | Content & Information | -0.478 |
| | Brother Cycles | Travel, Leisure & Sport | -0.488 |
| | LANDBAY | Finance & Payment | -0.493 |
| | EventsCase | SaaS/PaaS | -0.507 |
| | Brushlink | Healthcare | -0.533 |
| | Loyalzoo | SaaS/PaaS | -0.539 |
| | Storemates | Home &Personal | -0.581 |
| ↓ | HowNow | SaaS/PaaS | -0.590 |
| Least | Fund Ourselves | Finance & Payment | -0.609 |
| Successful | Glentham Capital | Finance & Payment | -0.607 |

SaaS: Software as a Service PaaS: Platform as a Service

Table 4.5 List of Company Performance

## 4.5 Conclusion

In the past decade or so, crowdfunding has rapidly gained importance as a method for firms to raise funds. Over the same period, there has been a rapid increase in the amount of research carried out with the objective of identifying the determinants of crowdfunding success. There is therefore no doubt that this is an interesting and important research area. As an extension of the previous chapter, equity crowdfunding is still the main analysis of this chapter. However, panel data collected from pitches of successfully funded equity crowdfunding campaigns is applied in this chapter. We have made a number of new contributions to this literature. First, while most previous studies have been based on cross-section data with a single observation per company on crowdfunding success, we have obtained multiple observations per company (in the form of "pitches") and accordingly applied panel-data methods to the resulting data set. The panel data demonstrates different pitches of the company as each company may raise funds for more than one time. Second, because truncation is quite likely to be ignored in research, especially when panel data is applied, we have taken account of the truncation in the data that arises because only successful pitches are observed. To accomplish this, we have constructed a new estimator, the "panel truncated regression estimator", and applied it to our panel data set. We find that the new estimator gives markedly different results from previously used estimators, suggesting that the correct choice of estimator is important. Third, we have applied the estimation procedure to a model specification that includes quadratic terms in all continuous explanatory variables. This has enabled us to deduce optimal values for each of these variables, that is, values predicted to maximise crowdfunding success. This information is clearly very useful to entrepreneurs setting out on crowdfunding campaigns. Finally, we have investigated the importance of the presence of particular words in crowdfunding announcements. We have found that certain words do indeed have a significant effect on success, some positive and some negative. This is a simple form of textual analysis. The interesting findings obtained here suggest that the use of more sophisticated methods of textual analysis in predicting crowdfunding success is a promising area for further research.

Since more individual investors would like to try to invest and equity crowdfunding would be a suitable choice for them to try, considering the factors analysed in this chapter probably would help them to have a better understanding before making decisions. However, because equity crowdfunding is different from conventional investment in the stock market, it still has a long way to go to develop into a mature system with more platforms doing this.

# Chapter 5

# Conclusion

## 5.1 Main Findings and Contributions

This thesis has been concerned with particular types of Cyber-Market. These markets have been in existence for only a relatively short period of time, and this is one of the things that attracts investors to them – the excitement factor. In addition to being interesting to the investment community and the general public, they are clearly also very interesting to the academic research community. This is clear from the rapid increase in the number of research papers on these topics in the very recent past.

Another reason for the increase in research activity is that new research methods have become available which facilitate this sort of research. The most important example of this is data-scraping. Data-scraping routines written in modern programming languages such as Python have automated the process of data collection, and have enabled access to datasets that it would not have been possible to assemble using traditional methods of data collection.

The thesis has attempted to contribute to this fast-growing literature. The Chapter 2 on Cryptocurrencies focused on the impact of investor sentiment on returns. The measures of investor sentiment are extracted from sources such as Google Trends, and here is another example of research possibilities opening up as a direct consequence of technological advances. It is hard to imagine how measures of investor sentiment could have been constructed before the Internet.

Our key findings in the Cryptocurrency chapter (Chapter 2) are summarised as follows. Our measure of stock market sentiment appeared to have no effect on the Bitcoin market, although it did have a negative effect on the gold market. The latter result is consistent with the popular idea of gold being a "safe haven". In contrast, our measure of economic conditions sentiment has a positive effect on the Bitcoin market, and no effect on gold. This result is quite striking because it suggests that cryptocurrencies are in some sense connected to the real economy.

Chapters 3 and 4 were concerned with the analysis of the crowdfunding market. An important fact about this market is that in addition to being very new, it is rapidly changing. One important aspect of this change is that the type of data that is available from crowdfunding platforms appears to have changed. In the early years, it appears that data was available on both successful and unsuccessful campaigns, and with this sort of data, the econometric investigation of the determinants of success is a relatively straightforward task. However, more recently (for some platforms at least), data is only made available on successful platforms. This data feature has been the principal theme of Chapters 3 and 4.

In Chapter 3, the data set analysed was a single cross-section, consisting of a single observation for each company. Accordingly, the econometric estimator used was the truncated regression model which is already established in the econometrics literature. In Chapter 4, the data set was panel data, since data on a sequence of "pitches" is available for each company. To analyse this data, we developed the panel truncated regression model and applied it for the first time. With each estimation, we have applied a formal test for truncation bias. The results from this test indicate that the bias from estimating models that disregard truncation is severe. Hence the importance of using estimators that allow for truncation is clear.

In Chapter 4, for the purpose of comparison, we reported results from simpler models as well as the panel truncated model. Here we observed that models that allow for the panel structure of the data but do not allow for truncation (e.g. the random effects linear model) produce estimates of effects that are biased towards zero, while models that allow for truncation but do not allow for the panel structure (e.g. cross-section truncated regression) produce estimates that are biased away from zero. This is useful in the sense that these two estimators, which are easy to estimate although we know they are not valid, may be useful in providing a lower and an upper bound to true effect sizes.

## 5.2 Recommendations for Future Research

Regarding Chapter 2 which was on the impact of sentiment on cryptocurrency returns, there are several possible avenues for future research. Most obviously, the data could be brought up-to-date. It would be particularly interesting to see the impact of the Covid-19 pandemic on the relationship between sentiment and returns. Second, while we focused on Bitcoin in Chapter 2 because it is a dominant cryptocurrency, it would be interesting to consider other cryptocurrencies. Third, while gold was used as a comparator in Chapter 2, there are other obvious candidates for comparators, such as major exchange rates.

In Chapter 2 the sentiment data were all collected from the internet they were all indices that have already been constructed and are directly available online. An alternative approach could be a lexicon-based sentiment analysis: from social media sites, posts relating to (e.g.) Bitcoin could be scraped and analysed. A lexicon dictionary could be used to assign a semantic orientation score to each word appearing in a post, and hence to assign a positive or negative sentiment score to the post. It could then be investigated whether information on the sentiment of posts is useful in predicting Bitcoin returns.

Another alternative approach is to investigate the effect of sentiment on volatility, rather than returns. Qadan and Nama (2018) used a GARCH model to analyse the effect of investor sentiment on oil price volatility. Similar methods could be applied to investigate the effects of sentiment on Bitcoin price volatility.

In terms of future research on ECF, our main recommendation must be: to wait for new data to appear. If these markets continue to become more popular and more well-known, they will continue to expand. One consequence of this will be that, in a few years, a typical data set is likely to be much larger than the ones used here. This will be very useful in terms of estimator precision. For example, in Chapter 4 we computed a set of optimal values of each of the continuous independent variables in the model, i.e. values that maximise the predicted level of success. These optimal values are clearly very useful but the confidence intervals were very wide. If this analysis is repeated in a few years' time, the larger data sets will give rise to narrower confidence intervals, and hence the advice to entrepreneurs will become more reliable.

A problem with the analysis of Chapter 4 is the possible endogeneity of the variable number of investors, which is treated as an independent variable in the model. The number of investors could clearly be used as an alternative measure of the success of the campaign, so it may be desirable to estimate a second model in which the number of investors is the dependent variable.

Another important suggestion for future research in the context of Crowdfunding is textual analysis. In Chapter 4, we performed a simple form of textual analysis, by simply including dummy variables for a set of words we suspected of being important in determining crowdfunding success. We found that some of these words are indeed important. This finding leads us to suggest the use of more sophisticated methods of textual analysis for the same purpose.

There were certain technical issues relating to the estimation of models that allowed for truncation, which lead to interesting areas of further research. For example, while the standard errors following the estimation of the panel truncated regression model were obtained using the Hessian matrix estimator, the robust (sandwich) covariance matrix estimator might be preferable, since it would allow for possible misspecifications in the model.

The link test was used as an omnibus test for misspecification. A great advantage of this test is that it is very easy to apply. The results from the link test sometimes appeared strange and contradictory, and the usefulness of this test in these sorts of models is a sensible avenue for further research. One welcome result from the link test, was that the full specification of the panel truncated regression model (which contained quadratic terms and word dummies, the final column of Table 4.3) was supported.

# Appendix A

# Cryptocurrency and PCA


Table A.1 Different Cryptocurrencies (VectorStock, 2019)

| Level | | | First Difference | | |
|---|---|---|---|---|---|
| variable | Test statistic | p-value I(0) | variable | Test statistic | p-value I(1) |
| $SVI_t^B$ | -3.316 | 0.0142 | $\Delta SVI_t^B$ | -20.701 | 0.0000 |
| $SVI_t^G$ | -7.163 | 0.0063 | $\Delta SVI_t^G$ | -17.565 | 0.0000 |
| $EPU_t$ | -8.515 | 0.0000 | - | - | - |
| $VIX_t$ | -6.552 | 0.0000 | - | - | - |
| $BBS_t$ | -7.252 | 0.0000 | - | - | - |
| $TED_t$ | -15.492 | 0.0085 | $\Delta TED_t$ | -27.050 | 0.0000 |
| $TY2R_t$ | 1.898 | 0.9985 | $\Delta TY2R_t$ | -13.558 | 0.0000 |
| $T10YR_t$ | -1.343 | 0.6091 | $\Delta TY10R_t$ | -13.413 | 0.0000 |
| $T30YR_t$ | -1.607 | 0.4800 | $\Delta TY30R_t$ | -13.687 | 0.0000 |
| $FF_t$ | 1.980 | 0.9986 | $\Delta FF_t$ | -19.776 | 0.0000 |
| $BI10YR_t$ | -1.198 | 0.6746 | $\Delta BI10YR_t$ | -13.297 | 0.0000 |
| $BAA_t$ | -1.632 | 0.4664 | $\Delta BAA_t$ | -13.886 | 0.0000 |

Table A.2 The results of the ADF test

| Lag | FPE | AIC | HQIC | SBIC |
|---|---|---|---|---|
| 0 | 0.013132 | 1.34305 | 1.39891 | 1.79505 |
| 1 | 0.013246 | 1.35165 | 1.42986 | 1.45331* |
| 2 | 0.012404* | 1.28599* | 1.38654* | 1.48193 |
| 3 | 0.012736 | 1.31231 | 1.4352 | 1.54608 |
| 4 | 0.013046 | 1.33633 | 1.48157 | 1.53597 |
| 5 | 0.013273 | 1.35348 | 1.52107 | 1.61783 |

Table A.3 VAR Lag Order Selection Criteria (model1 for Bitcoin)

| Lag | FPE | AIC | HQIC | SBIC |
|---|---|---|---|---|
| 0 | 0.006044 | 0.566975 | 0.636851 | 0.685987* |
| 1 | 0.006184 | 0.58998 | 0.62203* | 0.703924 |
| 2 | 0.005859* | 0.535915* | 0.667057 | 0.781709 |
| 3 | 0.006006 | 0.56069 | 0.635015 | 0.782423 |
| 4 | 0.006191 | 0.580542 | 0.681812 | 0.861979 |
| 5 | 0.006311 | 0.602685 | 0.728593 | 0.941092 |

Table A.4 VAR Lag Order Selection Criteria (model2 for Bitcoin)

| Lag | FPE | AIC | HQIC | SBIC |
|---|---|---|---|---|
| 0 | 0.156301 | 3.81978 | 3.83079 | 3.84717 |
| 1 | 0.078516 | 3.13129 | 3.16481 | 3.21462 |
| 2 | 0.076541* | 3.10582* | 3.13885* | 3.18799* |
| 3 | 0.07809 | 3.12585 | 3.1809 | 3.2628 |
| 4 | 0.079088 | 3.13853 | 3.21561 | 3.33026 |
| 5 | 0.079132 | 3.13907 | 3.23816 | 3.38557 |

Table A.5 VAR Lag Order Selection Criteria (model3 for Bitcoin)

| Lag | FPE | AIC | HQIC | SBIC |
|---|---|---|---|---|
| 0 | 0.000375 | -2.21216 | -2.20115 | -2.18477 |
| 1 | 0.000249 | -2.62036 | -2.58733 | -2.53819* |
| 2 | 0.000241* | -2.65566* | -2.60061* | -2.51871 |
| 3 | 0.000242 | -2.64928 | -2.57221 | -2.45756 |
| 4 | 0.000243 | -2.64541 | -2.54631 | -2.3989 |
| 5 | 0.000246 | -2.63373 | -2.5126 | -2.33244 |

Table A.6 VAR Lag Order Selection Criteria (model1 for gold)

| Lag | FPE | AIC | HQIC | SBIC |
|---|---|---|---|---|
| 0 | 0.000166 | -3.0258 | -3.01479 | -2.99841 |
| 1 | 0.000117 | -3.37872 | -3.34569 | -3.29655* |
| 2 | 0.000112* | -3.42047* | -3.36542* | -3.28353 |
| 3 | 0.000115 | -3.39372 | -3.31664 | -3.20199 |
| 4 | 0.000117 | -3.37846 | -3.27936 | -3.13195 |
| 5 | 0.000116 | -3.38599 | -3.26487 | -3.0847 |

Table A.7 VAR Lag Order Selection Criteria (model2 for gold)

| Lag | FPE | AIC | HQIC | SBIC |
|---|---|---|---|---|
| 0 | 0.002896 | -0.168713 | -0.157541 | -0.140938 |
| 1 | 0.001405 | -0.891918 | -0.858401* | -0.808594* |
| 2 | 0.00138* | -0.909671* | -0.853811 | -0.770798 |
| 3 | 0.0014 | -0.89544 | -0.817235 | -0.701017 |
| 4 | 0.001432 | -0.872751 | -0.772202 | -0.62278 |
| 5 | 0.001457 | -0.855693 | -0.7328 | -0.550172 |

Table A.8 VAR Lag Order Selection Criteria (model3 for gold)

| Component | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| Comp1 | 3.31515 | 1.90616 | 0.3014 | 0.3014 |
| Comp2 | 1.40899 | 0.294936 | 0.1281 | 0.4295 |
| Comp3 | 1.11405 | 0.082152 | 0.1013 | 0.5307 |
| Comp4 | 1.0319 | 0.0222503 | 0.0938 | 0.6246 |
| Comp5 | 1.00965 | 0.0621212 | 0.0918 | 0.7163 |
| Comp6 | 0.947528 | 0.174983 | 0.0861 | 0.8025 |
| Comp7 | 0.772546 | 0.0355714 | 0.0702 | 0.8727 |
| Comp8 | 0.736974 | 0.211679 | 0.0670 | 0.9397 |
| Comp9 | 0.525295 | 0.41538 | 0.0478 | 0.9875 |
| Comp10 | 0.109915 | 0.0819102 | 0.0100 | 0.9975 |
| Comp11 | .0280052 | - | 0.0025 | 1.0000 |

Principal components (eigenvectors)

| Variable | Comp1 | Comp2 | Comp3 | Comp4 | Comp5 | Comp6 | Comp7 | Comp8 | Comp9 | Comp10 | Comp11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Delta SVI_t^B$ | 0.0353 | 0.0262 | 0.1086 | -0.3180 | 0.7919 | 0.5076 | -0.0081 | 0.0022 | 0.0005 | 0.0050 | -0.0155 |
| $EPU_t$ | 0.0270 | -0.1340 | 0.6905 | -0.2917 | 0.0964 | -0.4712 | 0.3499 | 0.1553 | 0.2000 | 0.0387 | 0.0077 |
| $VIX_t$ | -0.0968 | 0.5023 | -0.4049 | -0.1499 | 0.0042 | -0.0263 | 0.5348 | 0.4804 | 0.1708 | -0.0722 | -0.0072 |
| $BBS_t$ | 0.1361 | -0.5099 | 0.0455 | 0.3966 | 0.0275 | 0.2129 | 0.1818 | 0.6368 | -0.2733 | 0.0515 | 0.0101 |
| $\Delta TED_t$ | 0.0061 | 0.3188 | 0.1760 | 0.7181 | 0.3146 | -0.0882 | 0.3572 | -0.3304 | -0.0826 | 0.0341 | 0.0133 |
| $\Delta T2YR_t$ | 0.4080 | 0.0307 | 0.0224 | 0.2299 | -0.0289 | 0.1539 | -0.1862 | 0.1127 | 0.7878 | 0.2204 | 0.1955 |
| $\Delta T10YR_t$ | 0.5377 | 0.0482 | 0.0134 | 0.0083 | -0.0243 | -0.0168 | 0.0010 | -0.0176 | 0.0247 | -0.4544 | -0.7072 |
| $\Delta T30YR_t$ | 0.5221 | 0.0696 | -0.0240 | -0.1137 | -0.0122 | -0.0608 | 0.0644 | -0.0639 | -0.2497 | -0.4297 | 0.6708 |
| $\Delta FF_t$ | -0.0202 | 0.2070 | 0.4309 | -0.0797 | -0.5104 | 0.6494 | 0.2592 | -0.1071 | -0.0598 | 0.0162 | -0.0013 |
| $\Delta BI10YR_t$ | 0.1219 | -0.5123 | -0.3544 | -0.1413 | -0.0343 | 0.0676 | 0.5698 | -0.4490 | 0.1595 | 0.1382 | -0.0223 |
| $\Delta BAA_t$ | 0.4762 | 0.2323 | -0.0439 | -0.1618 | -0.0226 | -0.1121 | -0.0141 | 0.0146 | -0.3698 | 0.7283 | -0.1027 |

Table A.9 PCA of $I_t^B$

| Component | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| Comp1 | 3.2557 | 2.05759 | 0.3617 | 0.3617 |
| Comp2 | 1.1981 | 0.151931 | 0.1331 | 0.4949 |
| Comp3 | 1.04617 | 0.0366375 | 0.1162 | 0.6111 |
| Comp4 | 1.00953 | 0.0552994 | 0.1122 | 0.7233 |
| Comp5 | 0.954235 | 0.149929 | 0.1060 | 0.8293 |
| Comp6 | 0.804306 | 0.217821 | 0.0894 | 0.9187 |
| Comp7 | 0.586485 | 0.469169 | 0.0652 | 0.9838 |
| Comp8 | 0.117316 | 0.0891648 | 0.0130 | 0.9969 |
| Comp9 | 0.0281509 | - | 0.0031 | 1.0000 |

Principal components (eigenvectors)

| Variable | Comp1 | Comp2 | Comp3 | Comp4 | Comp5 | Comp6 | Comp7 | Comp8 | Comp9 |
|---|---|---|---|---|---|---|---|---|---|
| $\Delta SVI_t^B$ | 0.0368 | 0.0216 | 0.3064 | 0.7875 | 0.5320 | 0.0276 | 0.0098 | 0.0042 | -0.0157 |
| $EPU_t$ | 0.0209 | 0.0998 | 0.7835 | 0.0667 | -0.5666 | 0.1604 | 0.1465 | 0.0549 | 0.0087 |
| $\Delta TED_t$ | 0.0115 | 0.5800 | -0.3925 | 0.3049 | -0.2764 | 0.5658 | -0.1376 | 0.0304 | 0.0125 |
| $\Delta T2YR_t$ | 0.4082 | 0.1114 | -0.1417 | -0.0318 | 0.0833 | 0.0213 | 0.8310 | 0.2517 | 0.1962 |
| $\Delta T10YR_t$ | 0.5427 | 0.0191 | -0.0007 | -0.0244 | -0.0206 | 0.0209 | 0.0359 | -0.4588 | -0.7013 |
| $\Delta T30YR_t$ | 0.5306 | -0.0334 | 0.0505 | -0.0095 | -0.0217 | -0.0063 | -0.2759 | -0.4266 | 0.6754 |
| $\Delta FF_t$ | -0.0162 | 0.4071 | 0.3370 | -0.5388 | 0.5561 | 0.3540 | -0.0778 | 0.0133 | -0.0019 |
| $\Delta BI10YR_t$ | 0.1078 | -0.6850 | -0.0304 | -0.0285 | 0.0425 | 0.6989 | -0.0331 | 0.1603 | -0.0212 |
| $\Delta BAA_t$ | 0.4934 | 0.0665 | 0.0347 | -0.0166 | -0.0232 | -0.1969 | -0.4294 | 0.7172 | -0.1122 |

Table A.10 PCA of $I_{Et}^B$

| Component | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| Comp1 | 3.31313 | 1.88777 | 0.3012 | 0.3012 |
| Comp2 | 1.42537 | 0.250785 | 0.1296 | 0.4308 |
| Comp3 | 1.17458 | 0.129278 | 0.1068 | 0.5376 |
| Comp4 | 1.0453 | 0.06235 | 0.0950 | 0.6326 |
| Comp5 | 0.982952 | 0.08582 | 0.0894 | 0.7219 |
| Comp6 | 0.897132 | 0.127688 | 0.0816 | 0.8035 |
| Comp7 | 0.769444 | 0.0406789 | 0.0699 | 0.8734 |
| Comp8 | 0.728766 | 0.203544 | 0.0663 | 0.9397 |
| Comp9 | 0.525221 | 0.415342 | 0.0477 | 0.9874 |
| Comp10 | 0.109879 | 0.0816544 | 0.0100 | 0.9974 |
| Comp11 | 0.0282247 | - | 0.0026 | 1.0000 |

Principal components (eigenvectors)

| Variable | Comp1 | Comp2 | Comp3 | Comp4 | Comp5 | Comp6 | Comp7 | Comp8 | Comp9 | Comp10 | Comp11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Delta SVI_t^G$ | 0.0196 | 0.2049 | 0.5003 | -0.2617 | 0.3162 | 0.6972 | -0.1175 | -0.1969 | 0.0131 | 0.0083 | 0.0038 |
| $EPU_t$ | 0.0274 | 0.1784 | 0.5424 | 0.1069 | -0.6771 | 0.0195 | 0.3373 | 0.2202 | 0.1978 | 0.0381 | 0.0070 |
| $VIX_t$ | -0.0974 | -0.50401 | -0.1881 | -0.3310 | 0.0946 | 0.3309 | 0.4350 | 0.4996 | 0.1702 | -0.0720 | -0.0072 |
| $BBS_t$ | 0.1370 | 0.5033 | -0.0543 | 0.3273 | 0.3268 | 0.0647 | 0.0913 | 0.6480 | -0.2757 | 0.0507 | 0.0100 |
| $\Delta TED_t$ | 0.0055 | -0.3246 | 0.0413 | 0.7480 | 0.1389 | 0.2605 | 0.3773 | -0.3099 | -0.0810 | 0.0346 | 0.0135 |
| $\Delta T2YR_t$ | 0.4081 | -0.0343 | 0.0019 | 0.1721 | 0.2125 | -0.0591 | -0.1949 | 0.0945 | 0.7875 | 0.2202 | 0.1957 |
| $\Delta T10YR_t$ | 0.5380 | -0.0486 | 0.0181 | 0.0045 | -0.0055 | -0.0051 | 0.0034 | -0.0169 | 0.0248 | -0.4535 | -0.7080 |
| $\Delta T30YR_t$ | 0.5221 | -0.0704 | -0 ∘ 0001 | -0.1003 | -0.0857 | 0.0098 | 0.0717 | -0.0558 | -0.2496 | -0.4305 | 0.6701 |
| $\Delta FF_t$ | -0.0193 | -0.1485 | 0.5392 | -0.1558 | 0.4840 | -0/5717 | 0.3118 | -0.0152 | -0.0622 | 0.0153 | -0.0014 |
| $\Delta BI10YR_t$ | 0.1222 | 0.4793 | -0.3522 | -0.2373 | 0.0580 | -0.0096 | 0.6259 | -0.3668 | 0.1897 | 0.1382 | -0.0221 |
| $\Delta BAA_t$ | 0.4761 | -0.2304 | 0.0154 | -0.1540 | -0.1277 | 0.0412 | -0.0196 | 0.0075 | -0.3697 | 0.7285 | -0.1021 |

Table A.11 PCA of $I_t^G$

| Component | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| Comp1 | 3.25304 | 2.05497 | 0.3614 | 0.3614 |
| Comp2 | 1.19806 | 0.0462501 | 0.1331 | 0.4946 |
| Comp3 | 1.15181 | 0.176907 | 0.1280 | 0.6225 |
| Comp4 | 0.974908 | 0.0851792 | 0.1083 | 0.7309 |
| Comp5 | 0.889728 | 0.0885513 | 0.0989 | 0.8297 |
| Comp6 | 0.801177 | 0.215429 | 0.0890 | 0.9187 |
| Comp7 | 0.585748 | 0.468594 | 0.0651 | 0.0939 |
| Comp8 | 0.117154 | 0.0887846 | 0.0130 | 0.9968 |
| Comp9 | 0.028369 | - | 0.0032 | 1.0000 |

Principal components (eigenvectors)

| Variable | Comp1 | Comp2 | Comp3 | Comp4 | Comp5 | Comp6 | Comp7 | Comp8 | Comp9 |
|---|---|---|---|---|---|---|---|---|---|
| $\Delta SVI_t^G$ | 0.0134 | -0.0245 | 0.6510 | -0.1356 | 0.7280 | -0.1570 | -0.0454 | 0.0144 | 0.0045 |
| $EPU_t$ | 0.0210 | 0.0835 | 0.5209 | 0.7613 | -0.2685 | 0.2066 | 0.1552 | 0.0536 | 0.0080 |
| $\Delta TED_t$ | 0.0111 | 0.5892 | -0.3345 | 0.2135 | 0.4595 | 0.5126 | -0.1419 | 0.0312 | 0.0128 |
| $\Delta T2YR_t$ | 0.4085 | 0.1142 | -0.0725 | -0.1352 | 0.0825 | 0.0043 | 0.8295 | 0.2511 | 0.1985 |
| $\Delta T10YR_t$ | 0.5432 | 0.0195 | 0.0084 | 0.0054 | 0.0042 | 0.0206 | 0.0358 | -0.4578 | -0.7022 |
| $\Delta T30YR_t$ | 0.5308 | -0.0340 | 0.0274 | 0.0343 | -0.0416 | 0.0016 | -0.2750 | -0.4274 | 0.6746 |
| $\Delta FF_t$ | -0.0156 | 0.3966 | 0.4321 | -0.5750 | -0.3975 | 0.4027 | -0.0686 | 0.0119 | -0.0021 |
| $\Delta BI10YR_t$ | 0.1081 | -0.6848 | -0.0101 | -0.0725 | 0.1141 | 0.6885 | -0.0300 | 0.1599 | -0.0209 |
| $\Delta BAA_t$ | 0.4936 | 0.0669 | 0.0048 | 0.0334 | -0.1821 | -0.1821 | -0.4279 | 0.7175 | -0.1115 |

Table A.12 PCA of $I_{E\,t}^{G}$

| Component | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| Comp1 | 1.216 | .431997 | 0.6080 | 0.6080 |
| Comp2 | 0.784002 | - | 0.3920 | 1.0000 |

Principal components (eigenvectors)

| Variable | Comp1 | Comp2 |
|---|---|---|
| $VIX_t$ | -0.7071 | 0.7071 |
| $BBS_t$ | 0.7071 | 0.7071 |

Table A.13 PCA of $I_{S_t}$

# Appendix B

# Equity Crowdfunding

Email from Seedrs indicating the data of unsuccessful campaigns are not available.

**Looking for Data**

JS Joel from Seedrs <joel.ippoliti@seedrs.com>
To: Xuerui Ma (ECO - Postgraduate Researcher)

Tue 01/12/2020 15:27

Hi Xuerui, Sorry, we cannot supply this information. For your purposes, if a company as funded once, you can assume it has had no unsuccessful campaigns. What you can't see are which businesses never funded in their first try. Good luck.
Support

**Joel** from Seedrs

On Tue, Dec 1, 2020 at 01:46 PM, "Xuerui Ma ( ECO - Postgraduate Researcher )" <xuerui.ma@uea.ac.uk> wrote:

> Dear Ines,
>
> Thanks again for your assistance with our queries.
>
> You say that information on unsuccessful pitches is not available to Seedrs users. We assume this is because the information is not of direct interest to investors.
>
> However, for our research questions to be properly investigated, we would require information on unsuccessful pitches (so that the data set is complete). We are now writing to enquire whether it might be possible for you to supply us with this data. We would quite understand if confidentiality or other issues prevent you from being able to do this.
>
> Best Regards,
>
> Xuerui Ma

Two tables below show the results of including the social media platform dummy variables to the model. These dummy variables show whether or not the campaign is on platform such as Facebook, Twitter, Instagram and LinkedIn. However unfortunately these variables do not have significant results on the level equity crowdfunding success.

Table B.1 Results from OLS and TRM. Dependent variable: $y_i$(log of Investment Amount) Model(3.9)

| VARIABLES | OLS | TRM |
|---|---|---|
| *equity* | 0.111*** | 0.123*** |
| | (0.00991) | (0.0127) |
| $equity^2$ | -0.00113*** | -0.00144*** |
| | (0.000268) | (0.000352) |
| $equity^3$ | 2.37e-06 | 4.23e-06* |
| | (1.69e-06) | (2.22e-06) |
| Facebook | -0.0329 | -0.0568 |
| | (0.0450) | (0.0529) |
| Twitter | 0.0490 | 0.0591 |
| | (0.0445) | ( 0.0517) |
| Instagram | -0.00287 | 0.0106 |
| | (0.0347) | (0.0418) |
| LinkedIn | 0.0329 | 0.0404 |
| | (0.0281) | (0.0331) |
| $log\,(valuation)$ | 1.021*** | 1.040*** |
| | ( 0.0113) | (0.0136) |
| Seedrs(base: Crowdcube) | 0.368*** | -0.443** |
| | (0.0697) | (0.436) |
| Constant | -3.475*** | -3.908*** |
| | (0.274) | ( 0.275) |

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table B.2 Results from OLS and TRM. Dependent variable: $ln(S_i)$ (log of Success Ratio) Model(3.13)

| VARIABLES | OLS | TRM |
|---|---|---|
| $equity$ | 0.110*** | 0.126*** |
| | (0.0100) | (0.0132) |
| $equity^2$ | -0.00114*** | -0.00354* |
| | (0.000268) | (0.000359) |
| $equity^3$ | 2.48e-06 | 4.05e-06* |
| | ( 1.70e-06) | (2.27e-06) |
| Facebook | -0.0313 | -.0643 |
| | ( 0.0451) | (0.0543) |
| Twitter | 0.0442 | 0.0737 |
| | (0.0451) | (0.0539) |
| Instagram | -0.00569 | 0.0208 |
| | ( 0.0350) | (0.0435871) |
| LinkedIn | 0.0320 | 0.0433 |
| | ( 0.0282) | (0.0339) |
| $log(InvestmentSought)$ | -0.979*** | -1.0525*** |
| | (0.0307) | (0.0381) |
| $log(valuation)$ | 1.00557*** | 1.0817*** |
| | ( 0.0263) | ( 0.0331) |
| Seedrs(base: Crowdcube) | -0.738*** | -9.373** |
| | (0.0543) | (0.743) |
| Constant | -3.476*** | -3.926*** |
| | ( 0.224) | (0.281) |

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Details of STATA do-file of Panel Truncated Regressionand Python codes for extracting information form Equity Crowdfunding Platform are shown in this Appendix.

```
gen date_s =mdy(month_s,day_s,year_s)
format date_s %td
gen date_c =mdy(month_c,day_c,year_c)
format date_c %td
gen date_i =mdy(month_i,day_i,year_i)

encode business, gen(business_id)

bysort business_id: gen t=_n

xtset business_id t

gen duration= date_c-date_s
gen company_age=date_s - date_i

encode sectors, gen(sectors_id)

*PANEL TRUNCATED REGRESSION

set more off

rename business_id i

**Read data

*use "ModifiedData1011.dta", clear

** make sure company number is named i, and pitch number t

*  DECLARE DATA TO BE PANEL DATA

gen heal = (health ==1)|(healthy ==1)

* Generate Threshold, c

gen double c=  0

* generate dependent variable (success)

gen double y=log( raised1/target)
```

```
gen double log_duration= log(duration)
gen double log_investors = log(investors)
*gen double log_teamnum =log(teamnum)
gen double log_equity = log(equityoffered)
*gen log_companyage=log(company_age)
gen double log_target=log(target)
*drop if y<=0.00
rename equityoffered equity
recast int teamnum

recast int s1-s14

recast double equity

drop if equity<=0

drop t
bysort i: gen int t=_n

xtset i t

** make sure all continuous variables appearing in model are in double precision

recast double y  log_target  log_duration log_equity teamnumber

* Censor problem variables

replace log_equity=0 if log_equity<0
replace log_target=10 if log_target<10

*heal organic planet quality data entertainment information

**make sure all integer variables are declared as integers:

recast int year_s

* generate quadratic terms:

gen double log_target_2= log_target^2

gen double log_duration_2= log_duration^2

gen double log_equity_2= log_equity^2
```

```
gen double teamnum_2= teamnum^2

* put explanatory variables in a list:

local list_explan " log_target log_target_2 log_duration log_duration_2 log_equity log_equity_2 teamnum teamnum_2 s5 s6 s8 h

* generate number of observations for each subject, T_i

by i: generate int T_i=_N

* NEED A NEW t INDEX

drop t

by i: gen int t=_n

order i t

*  DECLARE PANEL

xtset i t

* GENERATE INDICATOR VARIABLES FOR FIRST AND LAST OBSERVATION FOR EACH SUBJECT

by i: gen int first=1 if _n==1
by i: gen int last=1 if _n==_N

* APPEND (HORIZONTALLY) EACH SUBJECT'S FIRST ROW WITH 125 HALTON DRAWS
* (DIFFERENT BETWEEN SUBJECTS).  STORE NUMBER OF DRAWS AS "draws".

mat p=[3]
mdraws if first==1, neq(1) dr(125) prefix(h) primes(p)
scalar draws=r(n_draws)

*CREATE A VARIABLE LIST CONTAINING THE HALTON DRAWS
* ENSURE THEY ARE IN DOUBLE PRECISION
* COPY THE ROW OF HALTONS IN EACH BLOCK INTO ROWS 2-T OF SAME BLOCK

local hlist h1*

quietly{
foreach v of varlist `hlist' {
recast double `v'
by i: replace `v'=`v'[1] if `v'==.
```

```
replace `v'=invnorm(`v')
}
}

* LIKELIHOOD EVALUATION PROGRAM "my_rep" STARTS HERE:

capt prog drop my_rep
program define my_rep

* SPECIFY ARGUMENTS

args todo b lnppp g
tempvar zz1 zz2 xb p pp ppp upp uppp d1 d2 d3 dd1 dd2 dd3 ddd1 ddd2 ddd3 g1 g2 g3
tempname theta s_u_e
local hlist h1*

local y $ML_y1

* EXTRACT ELEMENTS OF PARAMETER VECTOR b

mleval `theta' =`b', eq(1) scalar
mleval `xb' = `b', eq(2)
mleval  `s_u_e' = `b', eq(3) scalar

* INITIALISE TEMPORARY VARIABLES

quietly gen double `zz1'=.
quietly gen double `zz2'=.

quietly gen double `p'=.
quietly gen double `pp'=.
quietly gen double `ppp'=0
quietly gen double `upp'=.
quietly gen double `uppp'=0

quietly gen double `d1'=.
quietly gen double `d2'=.
quietly gen double `d3'=.

quietly gen double `dd1'=.
quietly gen double `dd2'=.
quietly gen double `dd3'=.

quietly gen double `ddd1'=0
```

```
quietly gen double ‘ddd2’=0
quietly gen double ‘ddd3’=0

* LOOP FOR EVALUATION OF SUM (OVER Halton draws) OF PRODUCT (OVER t)
* pp AND ppp ARE FOR LIKELIHOOD FUNCTION;
* upp AND uppp ARE FOR NUMERATOR OF POSTERIOR RANDOM EFFECT FORMULA



quietly{
foreach v of varlist ‘hlist’ {

replace ‘zz1’=‘theta’*‘y’-‘xb’ - ‘s_u_e’*‘v’
replace ‘zz2’=‘theta’*c-‘xb’ - ‘s_u_e’*‘v’


replace  ‘p’= ‘theta’*normalden(‘zz1’)/(1-normal(‘zz2’))
by i: replace ‘pp’ = exp(sum(ln(‘p’)))
by i: replace ‘pp’=‘pp’[_N] if last~=1
replace ‘upp’=‘s_u_e’*‘v’*‘pp’
replace ‘ppp’=‘ppp’+‘pp’
replace ‘uppp’=‘uppp’+‘upp’

replace ‘d1’=(1/‘theta’)-‘zz1’*‘y’+normalden(‘zz2’)*c/(1-normal(‘zz2’))
replace ‘d2’=‘zz1’-normalden(‘zz2’)/(1-normal(‘zz2’))
replace ‘d3’=‘zz1’*‘v’-normalden(‘zz2’)*‘v’/(1-normal(‘zz2’))

by i: replace ‘dd1’ = sum(‘d1’)
by i: replace ‘dd2’ = ‘d2’
by i: replace ‘dd3’ = sum(‘d3’)

replace ‘ddd1’= ‘ddd1’ + ‘dd1’*‘pp’
replace ‘ddd2’= ‘ddd2’ + ‘dd2’*‘pp’
replace ‘ddd3’= ‘ddd3’ + ‘dd3’*‘pp’

}

* DIVISION BY Number of Draws TO GENERATE REQUIRED AVERAGES (OVER Haltons)
* COMPUTE POSTERIOR RANDOM EFFECT VARIABLE (u_post) AND SEND THIS TO MATA

quietly {
replace ‘ppp’=‘ppp’/draws

replace ‘uppp’=‘uppp’/draws
```

```
replace ppp=`ppp'
replace u_post=`uppp'/`ppp'

replace `ddd1'=`ddd1'/draws
replace `ddd2'=`ddd2'/draws
replace `ddd3'=`ddd3'/draws

by i: replace `ppp'=`ppp'[_N] if last~=1

}

* MLSUM COMMAND TO SPECIFY PER-SUBJECT LOG-LIKELIHOOD CONTRIBUTION

mlsum `lnppp'=ln(`ppp') if last==1

* MLVECSUM COMMAND TO COMPUTE ANALYTIC FIRST DERIVATIVES

mlvecsum `lnppp' `g1' = `ddd1'/`ppp' if last==1, eq(1)
mlvecsum `lnppp' `g2' = (`ddd2'/`ppp')  , eq(2)
mlvecsum `lnppp' `g3' = `ddd3'/`ppp' if last==1, eq(3)

mat `g' = (`g1',`g2',`g3')


replace g1=(`ddd1'/`ppp') if last==1
replace g2=(`ddd2'/`ppp')
replace g3=(`ddd3'/`ppp') if last==1

putmata  u_post ppp g1 g2 g3, replace

}
end

* "end" SIGNIFIES END OF LIKELIHOOD EVALUATION PROGRAM "my_rep"

* ESTIMATE SIMPLE PROBIT MODEL
* STORE ESTIMATES FROM SIMPLE PROBIT MODEL
* CREATE VECTOR OF STARTING VALUES (start) FOR PANEL PROBIT MODEL
* INITIALISE VARIABLE CONTAINING POSTERIOR RANDOM EFFECT (u_post)

truncreg y `list_explan',ll(0)
xtreg y `list_explan'
mat b_xtreg=e(b)
scalar sig_u_xtreg = (e(sigma_u))
```

```
scalar sig_e_xtreg = (e(sigma_e))

mat start = (1/sig_e_xtreg,b_xtreg/sig_e_xtreg,sig_u_xtreg/sig_e_xtreg)

* USE SUPERIOR STARTING VALUES!

*mat start=(1, .5, 1, -2.5, 1)

gen double u_post=.
gen double ppp=.
gen double g1=.
gen double g2=.
gen double g3=.

** constraints

 constraint 1 [xb]_b[log_equity]=1.0
  constraint 2 [xb]_b[c.log_equity#c.log_equity]=-0.11

    constraint 3 [xb]_b[log_equity]=1.381
   constraint 4 [xb]_b[c.log_equity#c.log_equity]=-0.18

ml model d1 my_rep /theta (xb: y  = `list_explan')  /s_u_e , constraints()
ml init start, copy
ml max , difficult  iter(20) trace gradient

** LOOK AT NUMBER OF OBSERVATIONS PER FIRM

tab T_i if last==1
hist  T_i if last==1

** recovery of structural form estimates

foreach v of varlist `list_explan' {
nlcom `v'_t: [xb]_b[`v']/_b[/theta]
}

nlcom cons_t: [xb]_b[_cons]/_b[/theta]

nlcom s_e: 1/_b[theta]

nlcom s_u: _b[s_u_e]/_b[theta]

** FIND OPTIMAL VALUES:
```

```
nlcom (target_star: exp(-[xb]_b[log_target]/(2*[xb]_b[log_target_2]))) (duration_star: exp(-[xb]_b[log_duration]/(2*[xb]

* EXTRACT POSTERIOR RANDOM EFFECT (u_post) GENERATED INSIDE EVALUATION PROGRAM

drop u_post

getmata u_post

sort u_post

* WHICH COMPANY HAS LOWEST AND HIGHEST POSTERIOR RANDOM EFFECT?

list i u_post if u_post!=.
```

```python
import requests
import fast
import math
import time
from openpyxl import Workbook
from pyquery import PyQuery as pq
import re

cookie = '_ga=GA1.2.570435255.1635301083; _gid=GA1.2.725921425.1635301083; user_attribution=seedrs; performance_cookie=true; functionality_and_profile_cookie=true; targeting_cookie=true; other_marketing


proxies = {
  "http": "http://127.0.0.1:25378/echo-pac?t=1435140",
  "https": "http://127.0.0.1:25378/echo-pac?t=1435140",
}

daili = 0

def getWords(text):
    text = text.replace("\n", " ")
    array=re.split('[ ,.]',text)
    #print(array)
    dic={}
    for i in array:
        k = i.strip()
        if k not in dic:
            dic[k] = 1
        else:
            dic[k] += 1
    return dic

def vget(page):
    global cookie,daili,proxies
    current_page = page - 1
    #https://www.seedrs.com/api/v1/businesses?per_page=20&expand[]=share_prices&sort=trending_desc&page=4&funded=true
    url = "https://www.seedrs.com/api/v1/businesses?per_page=50&page="+str(page)+"&sort=trending_desc&funded=true&available_shares=false&current_page="+str(current_page)+"&expand[]=share_prices"

    #print(url)

    payload={}
    headers = {

    'cookie': cookie,
    'referer': 'https://www.seedrs.com/invest/marketplace?sort=trending_desc&sort_collection=businesses',

    'X-SeedrsApplicationName': 'InvestorWeb',
    'X-SeedrsApplicationType': 'Investor',

    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/86.0.4240.198 Safari/537.36',

    }

    if daili==1:
        response = requests.request("GET", url, headers=headers, data=payload,proxies=proxies)
    else:
        response = requests.request("GET", url, headers=headers, data=payload)

    response.encode = "utf-8"
    return (response.text)
```

```python
def getTeam(url):
    global cookie,daili,proxies
    payload={}
    headers = {
    'cookie': cookie,
    'if-none-match': 'W/"a32d3ca3e96a4623339cc2123fc60283"',
    'sec-fetch-mode': 'navigate',
    'sec-fetch-site': 'none',
    'sec-fetch-user': '?1',
    'upgrade-insecure-requests': '1',
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/78.0.3904.108 Safari/537.36'
    }

    if daili==1:
        response = requests.request("GET", url, headers=headers, data=payload,proxies=proxies)
    else:
        response = requests.request("GET", url, headers=headers, data=payload)

    return (response.text)

def getMarket(url):
    global cookie,daili,proxies
    payload={}
    headers = {
    'cookie': cookie,
    'if-none-match': 'W/"0ac6839b0383f6d49ac62756975d042f"',
    'sec-fetch-mode': 'navigate',
    'sec-fetch-site': 'none',
    'sec-fetch-user': '?1',
    'upgrade-insecure-requests': '1',
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/78.0.3904.108 Safari/537.36'
    }

    if daili==1:
        response = requests.request("GET", url, headers=headers, data=payload,proxies=proxies)
    else:
        response = requests.request("GET", url, headers=headers, data=payload)

    return (response.text)

def getTarget(url):
    global cookie,daili,proxies
    payload={}
    headers = {
    'cookie': cookie,
    'if-none-match': 'W/"a66ae429dc0a544803c4050345bfcd56"',
    'referer': 'https://www.seedrs.com/businesses/revolut/sections/about',
    'sec-fetch-mode': 'navigate',
    'sec-fetch-site': 'none',
    'sec-fetch-user': '?1',
    'upgrade-insecure-requests': '1',
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/78.0.3904.108 Safari/537.36'
    }

    if daili==1:
        response = requests.request("GET", url, headers=headers, data=payload,proxies=proxies)
    else:
        response = requests.request("GET", url, headers=headers, data=payload)
```

```python
    return (response.text)

def getAbout(url):
    global cookie,daili,proxies
    payload={}
    headers = {
    'cookie': cookie,
    'if-none-match': 'W/"a66ae429dc0a544803c4050345bfcd56"',
    'referer': 'https://www.seedrs.com/businesses/revolut/sections/about',
    'sec-fetch-mode': 'navigate',
    'sec-fetch-site': 'none',
    'sec-fetch-user': '?1',
    'upgrade-insecure-requests': '1',
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/78.0.3904.108 Safari/537.36'
    }
    if daili==1:
        response = requests.request("GET", url, headers=headers, data=payload,proxies=proxies)
    else:
        response = requests.request("GET", url, headers=headers, data=payload)
    return (response.text)




fast.mkdir("./dat")
fast.mkdir("./html")

keywordArr = fast.txtGetLine("./keyword.txt", "gbk")

dataList = []


#541
count = 541

size = 50

pageNum = math.ceil(count/size)
page = 1

ts = ""

ki = 0

n = 0

#25


si = 1
ei = pageNum


#ei = 3
for page in range(si, ei+1):

    current_page = page - 1
```

```
fname = "./dat/"+str(page)+".json"


if fast.isFile(fname)==False:
    print(fname)
    print("get")
    s = vget(page)
    time.sleep(1.8)
    fast.toFile(fname, s)
    print(fname+"\t"+str(len(s)))

if fast.isFile(fname):
    arr = fast.getJson(fname)
    for r in arr['result']:
        n+=1
        msg = "n:"+str(n)+"/"+str(count)
        print(msg)


        name = r['slug']
        print("page:"+str(page)+"/"+str(ei)+" "+str(n)+"\t"+name)

        marketFile = "./html/market_"+str(name)+".html"
        if fast.isFile(marketFile):
            html = fast.getFile(marketFile)
        else:
            url = 'https://www.seedrs.com/businesses/{}/sections/market'.format(name)
            html = getMarket(url)
            time.sleep(0.1)
            print("\tget market "+url)
            fast.toFile(marketFile, html)

        aboutFile = "./html/about_"+str(name)+".html"
        if fast.isFile(aboutFile):
            aboutHtml = fast.getFile(aboutFile)
        else:
            aboutUrl = 'https://www.seedrs.com/businesses/{}/about'.format(name)
            aboutHtml = getAbout(aboutUrl)
            time.sleep(0.1)
            print("\tget about "+aboutUrl)
            fast.toFile(aboutFile, aboutHtml)

        targetList = []

        adoc = pq(aboutHtml)


        teamNum = 0

        for ar in adoc('.team .entrepreneur_container').items():
            teamNum+=1

        ai = 0
        ti = 0
        for ar in adoc('.PitchesList li').items():
            ai+=1
            if ai==1:
                continue
```

```python
ki+=1

link = ar('a').attr('href')


#https://www.seedrs.com/revolut/sections/team

teamUrl = 'https://www.seedrs.com{}/sections/team'.format(link)
#print(teamUrl)
ts+= teamUrl+"\n"

linkname = link.replace("/", "")
teamFile = "./html/nteam_"+str(linkname)+".html"
if fast.isFile(teamFile):
    teamHtml = fast.getFile(teamFile)
else:

    teamHtml = getTeam(teamUrl)
    time.sleep(0.1)
    print("\tget team "+teamUrl+"\t"+str(len(teamHtml)))
    fast.toFile(teamFile, teamHtml)


tdoc = pq(teamHtml)
teamStr = tdoc('#important-info').text()
mdate = fast.sm('as of*as', teamStr)

mdate = mdate.strip()


targetUrl = 'https://www.seedrs.com'+link


targetFile = "./html/target_"+str(name)+"_"+str(linkname)+".html"
if fast.isFile(targetFile):
    targetHtml = fast.getFile(targetFile)
else:
    targetHtml = getTarget(targetUrl)
    time.sleep(0.1)
    print("\tget target "+targetUrl)
    fast.toFile(targetFile, targetHtml)
#

target = fast.sm('<div class="investment_total_target">*</div>', targetHtml).replace("target", "").replace("\n", "")
#print(target)
#row['target'] =

vraise = fast.sm('<div class="CampaignProgress-text">*</div>', targetHtml).split('from')[0].strip()



t = fast.sm('<li class="ListGroup-row">*</li>', aboutHtml)
t = t.replace("\n", "")
tarr = t.split('</div>')

row = {}
try:
    row['IncorporationDate'] = fast.sm('Incorporation date*</b>', html).split('<b>')[1]
```

```
ti+=1

ki+=1

link = ar('a').attr('href')


#https://www.seedrs.com/revolut/sections/team

teamUrl = 'https://www.seedrs.com{}/sections/team'.format(link)
#print(teamUrl)
ts+= teamUrl+"\n"

linkname = link.replace("/", "")
teamFile = "./html/nteam_"+str(linkname)+".html"
if fast.isFile(teamFile):
    teamHtml = fast.getFile(teamFile)
else:

    teamHtml = getTeam(teamUrl)
    time.sleep(0.1)
    print("\tget team "+teamUrl+"\t"+str(len(teamHtml)))
    fast.toFile(teamFile, teamHtml)


tdoc = pq(teamHtml)
teamStr = tdoc('#important-info').text()
mdate = fast.sm('as of*as', teamStr)

mdate = mdate.strip()


targetUrl = 'https://www.seedrs.com'+link


targetFile = "./html/target_"+str(name)+"_"+str(linkname)+".html"
if fast.isFile(targetFile):
    targetHtml = fast.getFile(targetFile)
else:
    targetHtml = getTarget(targetUrl)
    time.sleep(0.1)
    print("\tget target "+targetUrl)
    fast.toFile(targetFile, targetHtml)
#

target = fast.sm('<div class="investment_total_target">*</div>', targetHtml).replace("target", "").replace("\n", "")
#print(target)
#row['target'] =

vraise = fast.sm('<div class="CampaignProgress-text">*</div>', targetHtml).split('from')[0].strip()



t = fast.sm('<li class="ListGroup-row">*</li>', aboutHtml)
t = t.replace("\n", "")
tarr = t.split('</div>')

row = {}
try:
```

```
    row['IncorporationDate'] = fast.sm('Incorporation date*</b>', html).split('<b>')[1]
except:
    row['IncorporationDate'] = ""

row['Pitchclosingdate'] = ar('.ListGroup-item:nth-child(1)').text()
try:
    row['Raised'] = ar('.ListGroup-item:nth-child(2)').text()
except:
    row['Raised'] = ""

row['Raised2'] = vraise

try:
    row['Pre-moneyValuation'] = ar('.ListGroup-item:nth-child(3)').text()
except:
    row['Pre-moneyValuation'] = ""

try:
    row['EquityOffered'] = ar('.ListGroup-item:nth-child(4)').text()
except:
    row['EquityOffered'] = ""

try:
    row['Investors'] = ar('.ListGroup-item:nth-child(5)').text()
except:
    row['Investors'] = ""

row['target'] = target

row['team date'] = mdate

sec = fast.sm('<th>Sectors</th>*</td>', html)
secArr = sec.split("</span>")
t = []
for sv in secArr:

    sv = fast.fm(sv)
    sv = sv.strip()

    sv = sv.replace("&amp;", "&")

    if len(sv)>0:
        t.append(sv)
secstr = "|".join(t)

row['Sectors'] = secstr

try:
    websiteHtml = fast.sm('<th>Social media</th>*<th>Sectors</th>', html).replace("\n", "")
except:
    websiteHtml = ""

try:
    row['Website'] = fast.fm(fast.sm('<th>Website</th><td>*</td>', websiteHtml)).strip()
except:
    row['Website'] = ""

dateArr = []
priceArr = []
```

```
    priceArr.append(price)


result = 0
targetDoc = pq(targetHtml)
ideaText = targetDoc("#detail_sections").text().strip()

#
dic = getWords(ideaText)

kwTotal = {}
for keyword in keywordArr:
    kwTotal[keyword] = 0

kwArr = []
for word in dic:
    if word in keywordArr:
        kwArr.append(word)
        msg = word +" find."
        #print(msg)
        kwTotal[word] = 1

#print("kwArr")
#print(kwArr)

#exit()
kwText = ""
if len(kwArr)>0:
    result = 1
    kwText = "|".join(kwArr)

rowData = [
    r['name'],
    row['IncorporationDate'] ,
    row['Pitchclosingdate'],
    row['target'],
    row['Raised'],
    row['Raised2'],
    row['Pre-moneyValuation'],
    row['EquityOffered'],
    row['Investors'],
    row['Sectors'],
    row['Website'],
    row['team date'],
    teamNum,
    teamStr,
    ideaText,
    result,
]

for v in kwTotal:
    num = kwTotal[v]
    rowData.append(num)

dataList.append(rowData)
```

```
#fast.toFile("ts.txt", ts)


wbname = "./Data.xlsx"
wb = Workbook()
ws = wb.active
titleArr = [
    "Business","Incorporation date","Pitch closing date","Target","Raised","Raised2","Pre-money valuation","Equity offered","Investors",
    "Sectors",
    "Website",
    "Team Date", "Team Num", "Important info",
    "ideaText",
    "Result"
]

for keyword in keywordArr:
    titleArr.append(keyword)

ws.append(titleArr)

ws.column_dimensions['A'].width = 25
ws.column_dimensions['B'].width = 25
ws.column_dimensions['C'].width = 25
ws.column_dimensions['D'].width = 20
ws.column_dimensions['E'].width = 20
ws.column_dimensions['F'].width = 20
ws.column_dimensions['G'].width = 20
ws.column_dimensions['H'].width = 20
ws.column_dimensions['I'].width = 12

ws.column_dimensions['J'].width = 30
ws.column_dimensions['K'].width = 20
ws.column_dimensions['L'].width = 15
ws.column_dimensions['M'].width = 12
ws.column_dimensions['N'].width = 50



for row in dataList:
    ws.append(row)

#saveexcel
wb.save(wbname)
print("OK")
```

# Bibliography


Abdi, H. and Williams, L. J. (2010), 'Principal component analysis', *Wiley interdisciplinary reviews: computational statistics* **2**(4), 433–459.

Abdullah, S. and Oseni, U. A. (2017), 'Towards a shari'ah compliant equity-based crowdfunding for the halal industry in malaysia', *International journal of business and society* **18**(S1), 223–240.

Agrawal, A., Catalini, C. and Goldfarb, A. (2014), 'Some simple economics of crowdfunding', *Innovation policy and the economy* **14**(1), 63–97.

Ahlers, G. K., Cumming, D., Günther, C. and Schweizer, D. (2015), 'Signaling in equity crowdfunding', *Entrepreneurship theory and practice* **39**(4), 955–980.

Aigner, D. J. and Hausman, J. A. (1980), 'Correcting for truncation bias in the analysis of experiments in time-of-day pricing of electricity', *The Bell Journal of Economics* pp. 131–142.

Andolfatto, D. (2013), 'Why gold and bitcoin make lousy money', *Economist's View* **24**.

Angerer, M., Brem, A., Kraus, S. and Peter, A. (2017), 'Start-up funding via equity crowdfunding in germany: A qualitative analysis of success factors', *The Journal of Entrepreneurial Finance (JEF)* **19**(1), 1–34.

Audretsch, D. B., Lehmann, E. E., Paleari, S. and Vismara, S. (2016), 'Entrepreneurial finance and technology transfer', *The Journal of Technology Transfer* **41**(1), 1–9.

Bachmann, A., Becker, A., Buerckner, D., Hilker, M., Kock, F., Lehmann, M., Tiburtius, P. and Funk, B. (2011), 'Online peer-to-peer lending-a literature review', *Journal of Internet Banking and Commerce* **16**(2), 1.

Baker, M. and Wurgler, J. (2006), 'Investor sentiment and the cross-section of stock returns', *The journal of Finance* **61**(4), 1645–1680.

Baker, S. R., Bloom, N. and Davis, S. J. (2016), 'Measuring economic policy uncertainty', *The quarterly journal of economics* **131**(4), 1593–1636.

Bank, B. B. (2014), 'Equity crowdfunding in the uk: Evidence from the equity tracker', *Hentet fra http://british-business-bank. co. uk/wpcontent/uploads/2015/03/230315-Equity-crowdfunding-report-final. pdf* .

Bapna, S. (2019), ‘Complementarity of signals in early-stage equity investment decisions: Evidence from a randomized field experiment’, *Management Science* **65**(2), 933–952.

Belleflamme, P., Lambert, T. and Schwienbacher, A. (2014), ‘Crowdfunding: Tapping the right crowd’, *Journal of business venturing* **29**(5), 585–609.

Block, J., Hornuf, L. and Moritz, A. (2018), ‘Which updates during an equity crowdfunding campaign increase crowd participation?’, *Small Business Economics* **50**(1), 3–27.

Borello, G., De Crescenzo, V. and Pichler, F. (2015), ‘The funding gap and the role of financial return crowdfunding: Some evidence from european platforms’, *The Journal of Internet Banking and Commerce* **20**(1), 1–20.

Bouoiyour, J. and Selmi, R. (2017), ‘The bitcoin price formation: Beyond the fundamental sources’, *arXiv preprint arXiv:1707.01284* .

Bradford, C. S. (2012), ‘Crowdfunding and the federal securities laws’, *Colum. Bus. L. Rev.* p. 1.

Braue, D. (2014), ‘Bitcoin confidence game is a ponzi scheme for the 21st century’.

Breen, R. et al. (1996), *Regression models: Censored, sample selected, or truncated data*, Vol. 111, Sage.

Brown, R., Mawson, S. and Rowe, A. (2019), ‘Start-ups, entrepreneurial networks and equity crowdfunding: A processual perspective’, *Industrial Marketing Management* **80**, 115–125.

Bruton, G., Khavul, S., Siegel, D. and Wright, M. (2015), ‘New financial alternatives in seeding entrepreneurship: Microfinance, crowdfunding, and peer–to–peer innovations’.

Bukovina, J., Martiček, M. et al. (2016), Sentiment and bitcoin volatility, Technical report, Mendel University in Brno, Faculty of Business and Economics.

Canette, I. (2016), Discrete-time survival analysis with stata, *in* ‘Stata Users Group Meeting Barcelona, October’.

Chaum, D. (1983), Blind signatures for untraceable payments, *in* ‘Advances in Cryptology: Proceedings of Crypto 82’, Springer, pp. 199–203.

Chen, H., Chong, T. T.-L. and Duan, X. (2010), ‘A principal-component approach to measuring investor sentiment’, *Quantitative Finance* **10**(4), 339–347.

Chohan, U. W. (2017), ‘A history of bitcoin’, *University of New South Wales Business School* .

Cicchiello, A. F. F. and Kazemikhasragh, A. (2022), ‘Tackling gender bias in equity crowdfunding: an exploratory study of investment behaviour of latin american investors’, *European Business Review* .

Clarke, N. (2018), ‘The crowd-funded book: an eighteenth-century revival’.
**URL:** *https://www.historyworkshop.org.uk/the-crowd-funded-book/*

Cleveland, W. S. (1979), ‘Robust locally weighted regression and smoothing scatterplots’, *Journal of the American statistical association* **74**(368), 829–836.

Coakley, J., Lazos, A. and Liñares-Zegarra, J. M. (2022), ‘Equity crowdfunding founder teams: Campaign success and venture failure’, *British Journal of Management* **33**(1), 286–305.

Coakley, J., Lazos, A. et al. (2021), ‘New developments in equity crowdfunding: A review’, *Review of Corporate Finance* **1**(3-4), 341–405.

Constancio, V. (2017), ‘Bitcoin is like tulipmania’, The Financial Times.

Cox, J. and Nguyen, T. (2017), ‘Does the crowd mean business? an analysis of rewards-based crowdfunding as a source of finance for start-ups and small businesses’, *Journal of Small Business and Enterprise Development* .

Cumming, D. J. and Hornuf, L. (2020), ‘Marketplace lending of smes’.

Cumming, D. J., Leboeuf, G. and Schwienbacher, A. (2020), ‘Crowdfunding models: Keep-it-all vs. all-or-nothing’, *Financial Management* **49**(2), 331–360.

Cumming, D., Meoli, M. and Vismara, S. (2019), ‘Investors’ choices between cash and voting rights: Evidence from dual-class equity crowdfunding’, *Research Policy* **48**(8), 103740.

Da, Z., Engelberg, J. and Gao, P. (2014), ‘The sum of all fears investor sentiment and asset prices’, *The Review of Financial Studies* **28**(1), 1–32.

Davis, J. (2011), ‘The crypto-currency: Bitcoin and its mysterious inventor.’.
**URL:** *https://www.newyorker.com/magazine/2011/10/10/the-crypto-currency*

Detrixhe, J. (2018), ‘Bitcoin trading in venezuela is skyrocketing amid 14,000% inflation.’, *Quartz* .

DeVries, P. D. (2016), ‘An analysis of cryptocurrency, bitcoin, and the future’, *International Journal of Business Management and Commerce* **1**(2), 1–9.

Di Pietro, F., Prencipe, A. and Majchrzak, A. (2018), ‘Crowd equity investors: An underutilized asset for open innovation in startups’, *California Management Review* **60**(2), 43–70.

Dilger, M. G., Jovanović, T. and Voigt, K.-I. (2017), ‘Upcrowding energy co-operatives–evaluating the potential of crowdfunding for business model innovation of energy co-operatives’, *Journal of environmental management* **198**, 50–62.

Dorfleitner, G., Hornuf, L. and Weber, M. (2017), ‘Dynamics of investor communication in equity crowdfunding. max planck insfitute for innovafion & compefifion research paper no. 17–06’.

Dority, B., Borchers, S. J. and Hayes, S. K. (2021), ‘Equity crowdfunding: Us title ii offerings using sentiment analysis’, *Studies in Economics and Finance* .

Dyhrberg, A. H. (2016), 'Bitcoin, gold and the dollar–a garch volatility analysis', *Finance Research Letters* **16**, 85–92.

Elendner, H., Trimborn, S., Ong, B. and Lee, T. M. (2017), The cross-section of crypto-currencies as financial assets: Investing in crypto-currencies beyond bitcoin, *in* 'Handbook of Blockchain, Digital Finance, and Inclusion, Volume 1', Elsevier, pp. 145–173.

Feibel, B. J. (2003), *Investment performance measurement*, John Wiley & Sons.

Fukuhara, M. et al. (2020), 'An overview of online alternative finance', *Public Policy Review* **16**(4), 1–19.

Gandal, N., Hamrick, J., Moore, T. and Oberman, T. (2018), 'Price manipulation in the bitcoin ecosystem', *Journal of Monetary Economics* **95**, 86–96.

Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D. and Giaglis, G. M. (2015), 'Using time-series and sentiment analysis to detect the determinants of bitcoin prices', *SSRN* .

Ghahramani, Z. and Jordan, M. I. (1995), 'Learning from incomplete data'.

Giga, A. (2017), 'Firm financing through equity crowdfunding', *Homerocket, nd* .

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M. C. and Siering, M. (2014), 'Bitcoin-asset or currency? revealing users' hidden intentions', *Revealing Users' Hidden Intentions (April 15, 2014). ECIS* .

Habla, Z. A. and Broby, D. (2019), 'Equity crowdfunding: literature review'.

Han, B. (2007), 'Investor sentiment and option prices', *The Review of Financial Studies* **21**(1), 387–414.

Hankin, A. (2019), 'Bitcoin pizza day: Celebrating the $80 million pizza order'.
**URL:** *https://www.investopedia.com/news/bitcoin-pizza-day-celebrating-20-million-pizza-order/*

Hartley, H. and Hocking, R. (1971), 'The analysis of incomplete data', *Biometrics* pp. 783–823.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica: Journal of the econometric society* pp. 1251–1271.

Hausman, J. A. and Wise, D. A. (1977), 'Social experimentation, truncated distributions, and efficient estimation', *Econometrica: Journal of the Econometric Society* pp. 919–938.

Henriques, I. and Sadorsky, P. (2018), 'Can bitcoin replace gold in an investment portfolio?', *Journal of Risk and Financial Management* **11**(3), 48.

Hervé, F. and Schwienbacher, A. (2018), 'Crowdfunding and innovation', *Journal of economic surveys* **32**(5), 1514–1530.

Herzenstein, M., Dholakia, U. M. and Andrews, R. L. (2011), 'Strategic herding behavior in peer-to-peer loan auctions', *Journal of Interactive Marketing* **25**(1), 27–36.

Hileman, G. and Rauchs, M. (2017), '2017 global blockchain benchmarking study', *University of Cambridge Judge Business School* .

Hoegen, A., Steininger, D. M. and Veit, D. (2018), 'How do investors decide? an interdisciplinary review of decision-making in crowdfunding', *Electronic Markets* **28**(3), 339–365.

Holmes, F. (2022), 'Decentralized assets like gold and bitcoin are looking more and more attractive'.
**URL:** *https://www.usfunds.com/resource/decentralized-assets-like-gold-and-bitcoin-are-looking-more-and-more-attractive/*

Honoré, B. E. and Powell, J. L. (1994), 'Pairwise difference estimators of censored and truncated regression models', *Journal of Econometrics* **64**(1-2), 241–278.

Hornuf, L. and Schmitt, M. (2016), 'Success and failure in equity crowdfunding', *CESifo DICE Report* **14**(2), 16–22.

Hornuf, L. and Schmitt, M. (2017), Equity crowdfunding in germany and the uk: Follow-up funding and firm survival, Technical report, CESifo working paper.

Hornuf, L., Schmitt, M. and Stenzhorn, E. (2018), 'Equity crowdfunding in germany and the united kingdom: Follow-up funding and firm failure', *Corporate Governance: An International Review* **26**(5), 331–354.

Hornuf, L. and Schwienbacher, A. (2015), 'Funding dynamics in crowdinvesting'.

Hornuf, L. and Schwienbacher, A. (2016), Crowdinvesting: angel investing for the masses?, *in* 'Handbook of research on business angels', Edward Elgar Publishing.

Hornuf, L. and Schwienbacher, A. (2018), 'Market mechanisms and funding dynamics in equity crowdfunding', *Journal of Corporate Finance* **50**, 556–574.

InterTradeIreland (2016), 'Equity crowdfunding resource'.

Johan, S. and Zhang, Y. (2021), 'Investors' industry preference in equity crowdfunding', *The Journal of Technology Transfer* pp. 1–29.

Kaiser, H. F. (1960), 'The application of electronic computers to factor analysis', *Educational and psychological measurement* **20**(1), 141–151.

Karalevicius, V., Degrande, N. and De Weerdt, J. (2018), 'Using sentiment analysis to predict interday bitcoin price movements', *The Journal of Risk Finance* **19**(1), 56–75.

Khavul, S., Chavez, H. and Bruton, G. D. (2013), 'When institutional change outruns the change agent: The contested terrain of entrepreneurial microfinance for those in poverty', *Journal of Business Venturing* **28**(1), 30–50.

Kim, H. and De Moor, L. (2017), ‘The case of crowdfunding in financial inclusion: A survey’, *Strategic Change* **26**(2), 193–212.

Kleemann, F., Voß, G. G. and Rieder, K. (2008), ‘Un (der) paid innovators: The commercial utilization of consumer work through crowdsourcing’, *Science, technology & innovation studies* **4**(1), 5–26.

Klein, J. P. and Moeschberger, M. L. (2003), *Survival analysis: techniques for censored and truncated data*, Vol. 2, Springer.

Klein, T., Thu, H. P. and Walther, T. (2018), ‘Bitcoin is not the new gold–a comparison of volatility, correlation, and portfolio performance’, *International Review of Financial Analysis* **59**, 105–116.

Kleinert, S. and Mochkabadi, K. (2021), ‘Gender stereotypes in equity crowdfunding: the effect of gender bias on the interpretation of quality signals’, *The Journal of Technology Transfer* pp. 1–22.

Kliber, A., Marszałek, P., Musiałkowska, I. and Świerczyńska, K. (2019), ‘Bitcoin: Safe haven, hedge or diversifier? perception of bitcoin in the context of a country’s economic situation—a stochastic volatility approach’, *Physica A: Statistical Mechanics and its Applications* **524**, 246–257.

Klöhn, L. and Hornuf, L. (2012), ‘Crowdinvesting in deutschland’, *Zeitschrift für Bankrecht und Bankwirtschaft* **24**(4), 237–266.

Konhäusner, P., Thielmann, M., Câmpian, V. and Dabija, D.-C. (2021), ‘Crowdfunding for independent print media: E-commerce, marketing, and business development’, *Sustainability* **13**(19), 11100.

Kräussl, R., Lehnert, T. and Rinne, K. (2017), ‘The search for yield: implications to alternative investments’, *Journal of Empirical Finance* **44**.

Kumar, P., Langberg, N. and Zvilichovsky, D. (2016), ‘(crowd) funding innovation: Financing constraints, price discrimination and welfare’, *Price Discrimination and Welfare (June 5, 2016)* .

Kuppuswamy, V. and Bayus, B. L. (2017), ‘Does my contribution to your crowdfunding project matter?’, *Journal of Business Venturing* **32**(1), 72–89.

Lee, C. K., Brown, C., Gralla, R. J., Hirsh, V., Thongprasert, S., Tsai, C.-M., Tan, E. H., Ho, J. C.-M., Chu, D. T., Zaatar, A. et al. (2013), ‘Impact of egfr inhibitor in non–small cell lung cancer on progression-free and overall survival: a meta-analysis’, *Journal of the National Cancer Institute* **105**(9), 595–605.

Lee, T. B. (2013), ‘Bitcoin needs to scale by a factor of 1000 to compete with visa. here’s how to do it’, *The Washington Post* .

Li, J. (2016), ‘Equity crowdfunding in china: Current practice and important legal issues’, *Asian Bus. Law.* **18**, 59.

Lin, M., Prabhala, N. R. and Viswanathan, S. (2013), 'Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending', *Management science* **59**(1), 17–35.

Lukkarinen, A. and Schwienbacher, A. (2020), 'Secondary market listings in equity crowdfunding: The missing link?', *Available at SSRN 3725498* .

Lukkarinen, A., Teich, J. E., Wallenius, H. and Wallenius, J. (2016), 'Success drivers of online equity crowdfunding campaigns', *Decision Support Systems* **87**, 26–38.

Maddala, G. S. (1983), *Limited-dependent and qualitative variables in econometrics*, number 3, Cambridge university press.

Malkiel, B. G. (1989), 'Is the stock market efficient?', *Science* **243**(4896), 1313–1318.

Mandel, M. (2007), 'Censoring and truncation—highlighting the differences', *The American Statistician* **61**(4), 321–324.

Manual, R. G. (1989), 'Parametric bayesian estimation of age-specific survival for left-truncated and right-censored capture-recapture/recovery data.'.

Manyika, J. and Roxburgh, C. (2011), 'The great transformer: The impact of the internet on economic growth and prosperity', *McKinsey Global Institute* **1**(0360-8581).

Masters, T. (2013), 'Marillion 'understood where the internet was going early on''.
**URL:** *https://www.bbc.co.uk/news/entertainment-arts-23881382*

Miglo, A. (2016), 'Crowdfunding: Balancing imperfect informafion and moral hazard considerafions'.

Mochkabadi, K. and Volkmann, C. K. (2020), 'Equity crowdfunding: a systematic review of the literature', *Small Business Economics* **54**(1), 75–118.

Mokhtarrudin, A., Masrurah, I. and Muhamad, S. (2017), 'Crowdfunding as a funding opportunity for youth start-ups in malaysia.', *Pertanika Journal of Social Sciences & Humanities* **25**.

Mollick, E. (2014), 'The dynamics of crowdfunding: An exploratory study', *Journal of business venturing* **29**(1), 1–16.

Moreno, E. C. (2016), 'Bitcoin in argentina: inflation, currency restrictions, and the rise of cryptocurrency'.

Nakamoto, S. (2008), 'Bitcoin: A peer-to-peer electronic cash system'.

Narayanan, A., Bonneau, J., Felten, E., Miller, A. and Goldfeder, S. (2016), *Bitcoin and cryptocurrency technologies: a comprehensive introduction*, Princeton University Press.

Nascimento, M. and Querette, E. (2013), Crowdfunding: challenges and alternatives to financing the innovative entrepreneurship in brazil, *in* '3rd Conference of the International

Consortium of Studies on Innovation and Entrepreneurship, November, Rio de Janeiro, Brazil'.

Nevin, S., Gleasure, R., O'Reilly, P., Feller, J., Li, S. and Christoforo, J. (2017), Large crowds or large investments?: How social identity influences the commitment of the crowd, *in* 'ECIS 2017 Proceedings', Association for Information Systems. AIS Electronic Library (AISeL), pp. 2802–2813.

Nitani, M. and Riding, A. (2017), On crowdfunding success: Firm and owner attributes and social networking, *in* '2017 Emerging Trends in Entrepreneurial Finance Conference'.

Nutting, M. R. and Freedman, D. M. (2015), *Equity crowdfunding for investors: A guide to risks, returns, regulations, funding portals, due diligence, and deal terms*, John Wiley & Sons.

Oehlert, G. W. (1992), 'A note on the delta method', *The American Statistician* **46**(1), 27–29.

Olsen, R. J. (1978), 'Note on the uniqueness of the maximum likelihood estimator for the tobit model', *Econometrica: Journal of the Econometric Society* pp. 1211–1215.

Patil, A. (2020), 'Principal component analysis (pca)'.
**URL:** *https://medium.com/analytics-vidhya/principal-component-analysis-pca-8a0fcba2e30c*

Pelizzon, L., Riedel, M. and Tasca, P. (2016), 'Classification of crowdfunding in the financial system', *Banking Beyond Banks and Money* pp. 5–16.

Perez, S. (2020), 'Paypal says all users in us can now buy, hold and sell cryptocurrencies'.
**URL:** *https://techcrunch.com/2020/11/12/paypal-says-all-users-in-u-s-can-now-buy-hold-and-sell-cryptocurrencies/?*

Peters, S. (2000), 'On the use of the reset test in microeconometric models', *Applied Economics Letters* **7**(6), 361–365.

Piva, E. and Rossi-Lamastra, C. (2018), 'Human capital signals and entrepreneurs' success in equity crowdfunding', *Small Business Economics* **51**(3), 667–686.

Pollack, J. M., Maula, M., Allison, T. H., Renko, M. and Günther, C. C. (2021), 'Making a contribution to entrepreneurship research by studying crowd-funded entrepreneurial opportunities'.

Preqin (2022), 'Future of alternatives 2027 press release'.
**URL:** *https://www.preqin.com/Portals/0/Documents/FOA-20press-20release-20FINAL.pdf?ver=2022-10-05-093626-147*

Qadan, M. and Nama, H. (2018), 'Investor sentiment and the price of oil', *Energy Economics* **69**, 42–58.

Ralcheva, A. and Roosenboom, P. (2020), 'Forecasting success in equity crowdfunding', *Small Business Economics* **55**(1), 39–56.

Ramsey, J. B. (1969), ‘Tests for specification errors in classical linear least-squares regression analysis’, *Journal of the Royal Statistical Society: Series B (Methodological)* **31**(2), 350–371.

Rossi, A., Vanacker, T. R. and Vismara, S. (2020), ‘Equity crowdfunding: New evidence from us and uk markets’, *Available at SSRN* .

Seedrs (2021), ‘What is the seedrs secondary market?’.
**URL:** *https://www.seedrs.com/insights/our-news/what-is-the-seedrs-secondary-market*

Sharma, D., Verma, R. and Sam, S. (2021), ‘Adoption of cryptocurrency: an international perspective’, *International Journal of Technology Transfer and Commercialisation* **18**(3), 247–260.

Shu, H.-C. (2010), ‘Investor mood and financial markets’, *Journal of Economic Behavior & Organization* **76**(2), 267–282.

Signori, A. and Vismara, S. (2016), ‘Returns on investments in equity crowdfunding’, *Available at SSRN 2765488* .

Son Turan, S. (2015), ‘Stakeholders in equity-based crowdfunding: respective risks over the equity crowdfunding lifecycle’, *Journal of Financial Innovation* .

StataCorp, L. (2021), ‘Stata statistical software: release 17 college station’, *TX StataCorp LLC* .

Statista (2017), ‘Equity based crowdfunding transaction value in europe. ed. statista.’.

Statista (2021), ‘Overall cryptocurrency market capitalization per week from july 2010 to december 2021’.
**URL:** *https://www.statista.com/statistics/730876/cryptocurrency-maket-value/*

Stekli, J. and Cali, U. (2020), ‘Potential impacts of blockchain based equity crowdfunding on the economic feasibility of offshore wind energy investments’, *Journal of Renewable and Sustainable Energy* **12**(5), 053307.

Stevenson, R. M., Kuratko, D. F. and Eutsler, J. (2019), ‘Unleashing main street entrepreneurship: Crowdfunding, venture capital, and the democratization of new venture investments’, *Small Business Economics* **52**(2), 375–393.

TheWallStreetJournal (2021), ‘Sandeep sood: The rise of crowdfunding’.
**URL:** *https://www.wsj.com/articles/BL-232B-3096*

Train, K. E. (2009), *Discrete choice methods with simulation*, Cambridge university press.

Trussell, J. and Bloom, D. E. (1979), ‘A model distribution of height or weight at a given age’, *Human Biology* pp. 523–536.

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

VectorStock (2019), 'Cryptocurrency types icons set flat style vector image'. **URL:** *https://www.vectorstock.com/royalty-free-vector/cryptocurrency-types-icons-set-flat-style-vector-20435236*

Vismara, S. (2016), 'Equity retention and social network theory in equity crowdfunding', *Small Business Economics* **46**(4), 579–590.

Vismara, S. (2021), 'Expanding corporate finance perspectives to equity crowdfunding', *The Journal of Technology Transfer* pp. 1–11.

Vulkan, N., Åstebro, T. and Sierra, M. F. (2016), 'Equity crowdfunding: A new phenomena', *Journal of Business Venturing Insights* **5**, 37–49.

Walthoff-Borm, X., Schwienbacher, A. and Vanacker, T. (2018), 'Equity crowdfunding: First resort or last resort?', *Journal of Business Venturing* **33**(4), 513–533.

Whaley, R. E. (2008), 'Understanding vix', *SSRN* .

Whiteley, S. and Rambarran, S. (2016), *The Oxford handbook of music and virtuality*, Oxford University Press.

Williams, D., Liao, X., Xue, Y., Carin, L. and Krishnapuram, B. (2007), 'On classification with incomplete data', *IEEE transactions on pattern analysis and machine intelligence* **29**(3), 427–436.

Yasar, B. (2021), 'The new investment landscape: Equity crowdfunding', *Central Bank Review* **21**(1), 1–16.

Zargar, F. N. and Kumar, D. (2019), 'Informational inefficiency of bitcoin: A study based on high-frequency data', *Research in International Business and Finance* **47**, 344–353.

Zhu, P., Zhang, X., Wu, Y., Zheng, H. and Zhang, Y. (2021), 'Investor attention and cryptocurrency: Evidence from the bitcoin market', *PLoS One* **16**(2), e0246331.