

CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Bachelor's thesis

2019

Tereza Navarová

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies



Tereza Navarová

**Crowdfunding:
What makes a project desirable?**

Bachelor's thesis

Prague 2019

Declaration of Authorship

I hereby declare that I compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

I grant to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agree with the thesis being used for study and scientific purposes.

Prague, July 29, 2019

Tereza Navarová

Acknowledgments

I would like to thank my supervisor, doc. PhDr. Julie Chytilová, Ph.D., for her time, guidance and valuable suggestions for this thesis. Most importantly, I wish to thank my family and friends for their support and encouragement throughout my whole studies.

Bibliographic note

NAVAROVÁ, Tereza. *Crowdfunding: What makes a project desirable?*. 43 p. Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies. Supervisor doc. PhDr. Julie Chytilová, Ph.D.

Abstract

Crowdfunding has become a popular option for financing new projects and businesses. The study explores the factors that influence the project's success from the creator's point of view. The analysis of a cross-sectional dataset from one of the most famous czech reward-based crowdfunding platforms HitHit is used to confirm or reject the hypotheses of existing studies from the global scale for the case of the Czech Republic. First, a logistic regression is applied to estimate the effects of the individual factors on the probability of a project being successfully funded. Secondly, the OLS regression is introduced to estimate the effects on the overall percentage of funding of a project in place of the binary variable in the logit model. We find a significant positive relationship between the success and the following factors: a video presence, a creator's location within a large city, a shorter campaign duration, a number of rewards and updates during the campaign, and projects falling into the category *education*. Several suggestions on future research are presented.

Keywords crowdfunding, P2P, OLS regression, logit, shared economy

Author's e-mail navarova.terez@gmail.com

Supervisor's e-mail julie.chytilova@cuni.cz

Abstrakt

Crowdfunding se stává čím dál oblíbenější formou financování nových projektů. Tato práce se zaměřuje na faktory ovlivňující úspěšnost projektu z pohledu zakladatele. Analýzou poskytnutých dat jednou z nejznámějších českých crowdfundingových platform potvrdíme či odmítáme hypotézy, které jsou zaběhnuté ve studiích zaměřujících se na odměnový crowdfunding v celosvětovém měřítku pro případ České republiky. Nejprve byla použita logistická regrese k zjištění faktorů, které ovlivňují pravděpodobnost úspěšného zafinancování projektu. Dále používáme metodu nejmenších čtverců a jako nezávislou proměnnou volíme celkové procento zafinancování namísto stručnější binární proměnné v logit modelu. Nalézáme signifikantně pozitivní vztah mezi úspěchem projektu a následujícími faktory: prezencí videa, zakladatelovou lokací ve větším městě, kratší dobou trvání kampaně, počtem odměn a aktualit a projekty v kategorii vzdělání. V závěru nabízíme další možnosti pro budoucí výzkum.

Klíčová slova

crowdfunding, P2P, metoda nejmenších čtverců, logit, sdílená ekonomika

E-mail autora

navarova.tereza@gmail.com

E-mail vedoucího práce

julie.chytilova@cuni.cz

Bachelor's Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague



Author's name and surname: Tereza Navarová

E-mail: navarova.tereza@gmail.com

Phone: +420 720 153 054

Supervisor's name: doc. PhDr. Julie Chytilová, Ph.D.

Supervisor's email: julie.chytilova@fsv.cuni.cz

Notes: Please enter the information from the proposal to the Student Information System (SIS) and submit the proposal signed by yourself and by the supervisor to the Academic Director ("garant") of the undergraduate program.

Proposed Topic:

Crowdfunding in the Czech Republic: What makes a project desirable?

Preliminary scope of work:

Research question and motivation

In my thesis I would like to focus on a rather new approach to alternative financing; reward-based crowdfunding. Reward-based crowdfunding found its way in the Czech Republic in 2011 provided by Hihit.cz, the first Czech crowdfunding platform. There are other types of crowdfunding platforms, but the reward-based is the most common and also the most accessible to a common citizen. My research question would be: What are the main factors that influence the success rate of a crowdfunded project?

The idea is to analyse the success rate in terms of characteristics of a project instead of analysing it from the other point of view, as of the characteristics of the funders. Even though crowdfunding has been active for a few years now, there is still a relatively small number of research papers on this topic.

Contribution

My thesis could serve as a help to navigate starting entrepreneurs who are thinking about financing their projects through crowdfunding in analysing their possibilities of succession when setting up a campaign as well as an informative analysis for crowdfunding platforms in the Czech Republic.

Methodology

The data I will be working with will be provided by the Hihit platform on all successfully and unsuccessfully completed projects and with the help of basic econometric models I will study the influence of different factors (video presence, update frequency, target amount, type of reward, etc.) on the success rate of a campaign which will be measured by how many pp has the project exceeded the target amount and how quickly it reached the target.

Outline

- 1) Introduction
- 2) Crowdfunding background; both worldwide and in the Czech Republic
- 3) Literature review
- 4) Data analysis and empirical model
- 5) Results
- 6) Conclusion

List of academic literature:

Bibliography

- Armin, Paul Belle, Thomas Lambert, and Armin Schwienbacher. 2011. “‘ Crowdfunding : Tapping the Right Crowd ’ Référéncé Bibliographique Journal of Business Venturing Crowdfunding : Tapping the Right Crowd ☆.” 29(5): 585–609.
- Giudici, Giancarlo, Massimiliano Guerini, and Cristina Rossi-Lamastra. 2018. “Reward-Based Crowdfunding of Entrepreneurial Projects: The Effect of Local Altruism and Localized Social Capital on Proponents’ Success.” *Small Business Economics* 50(2): 307–24.
- Makýšová, Lucia, and Gabriela Vaceková. 2017. “Profitable Nonprofits? Reward-Based Crowdfunding in the Czech Republic.” *NISPAcee Journal of Public Administration and Policy* 10(2): 203–27.
- Mollick, Ethan. 2014. “The Dynamics of Crowdfunding: An Exploratory Study.” *Journal of Business Venturing* 29(1): 1–16. <http://dx.doi.org/10.1016/j.jbusvent.2013.06.005>.
- Šoltés, Michal, and Tomáš Štofa. 2016. “Crowdfunding – The Case of Slovakia and the Czech Republic.” *Quality Innovation Prosperity* 20(2): 89–104.

Contents

List of Tables	ix
List of Figures	x
Acronyms	xi
1 Introduction	1
2 Crowdfunding	4
2.1 Definitions	4
2.2 Taxonomy	5
2.2.1 Financial Crowdfunding	5
2.2.2 Non-financial Crowdfunding	6
2.3 Economic context	7
2.4 Crowdfunding Campaign Phases	9
3 Literature Review	13
3.1 Factors influenceable by the creator	13
3.2 Other influential factors	16
4 Data	17
4.1 Data selection	17
4.2 Variables and Hypotheses	18
4.3 Descriptive statistics	22
5 Methodology	26
5.1 Logit model	26
5.2 OLS model	29
6 Results	31
6.1 Regression results - the logit model	31

6.1.1	Interpretation of results	34
6.2	Regression results - the OLS model	38
6.2.1	Interpretation of results	40
6.3	Causality vs Correlation	41
7	Conclusion	42
	Bibliography	44
A	Tests and Assumptions	I
B	Additional Regressions	III

List of Tables

4.1	Dependent variables	18
4.2	Independent variables	19
4.3	Descriptive statistics/dependent variables	22
4.4	Descriptive statistics/independent variables	24
6.1	Likelihood ratio test results	32
6.2	Logistic regression	33
6.3	Average Partial Effects	35
6.4	OLS regression with robust errors	39
A.1	LPM model - robustness check for logit	I
A.2	Variance Inflation Factors	II
B.1	OLS regression - only successful projects	III
B.2	OLS regression - <i>average_contribution</i>	IV

List of Figures

2.1	Online Alternative Finance Volume by Country in 2017 (€millions)	8
2.2	Success rates of crowdfunding platforms	11
4.1	Funding levels of unsuccessful projects	23
4.2	Funding levels of successful projects	23

Acronyms

AON all-or-nothing

APE average partial effect

EU European Union

KIA keep-it-all

OLS ordinary least squares

P2P peer-to-peer

PEA partial effect at average

UK United Kingdom

US United States

Chapter 1

Introduction

Crowdfunding, as a concept, has been known for a very long time. As an example, we can mention the case of Alexander Pope, who used the concept of crowdfunding in order to publish his English translation of *The Iliad*. Before starting the work, he managed to assemble as many as 750 people who were willing to pay a certain amount of money for the yet unpublished books of *The Iliad*. As a reward, he offered a copy of each of the volumes to the contributors.

However, crowdfunding, as we know it nowadays, a tool to finance projects of starting entrepreneurs by other people via the Internet, referred to as *the crowd*, has only appeared in the 21st century. The main milestone for crowdfunding was the year 2009 when many entrepreneurs faced difficulties obtaining funds from bank loans or attracting potential investors after the financial crisis (Belleflamme et al. 2014). Therefore, they had to look for alternative sources of funding.

The term crowdfunding is closely related to *crowdsourcing*. It is a type of activity where people on the Internet voluntarily participate in a task based on an open call of another person or entity. To meet the assumptions of the definition, there should always be a mutual benefit both to the volunteers and the organization behind the task. To list some of its cases; city governments often use crowdsourcing as a way to collect information on how to improve their communities, organizations seek potential labourers online as well as money to finance their projects. This concretely is the case of crowdfunding (Brabham 2013).

Crowdfunding is another type of financial funding. The idea is that a large number of individuals decides to pay a tiny amount of money in order to help entrepreneurs to raise funding, rather than a small number of individuals de-

ceding to invest relatively large amounts of money in a project (Belleflamme et al. 2014). A vital aspect of crowdfunding is the benefits the backers enjoy by engaging in the process.

To this day, Kickstarter¹, the largest and the most popular crowdfunding platform in the world, has raised more than \$4,174,490,000 with almost 160,000 successfully funded projects. This platform, operating in English, has a global impact and its backers come from all around the globe. However, in the Czech Republic, the most well-known platform *HitHit*², has raised over 129,860,000 CZK (approximately \$5,722,212)³ with around 1,000 successfully funded projects since its beginnings in 2012.

Even though crowdfunding is becoming a more and more attractive topic for researchers, the case of the Czech Republic remains to some extent unexplored. Therefore, more research is still needed.

The objective of my thesis is to complete the world literature concerning reward-based crowdfunding with the case of the Czech Republic. While there have been several studies focusing on crowdfunding in the Czech Republic, many of them focused only on particular categories or types of crowdfunding. Several studies covered the donation-based crowdfunding or reward-based crowdfunding of NGOs. A comprehensive research of rewards-based crowdfunding in the Czech Republic has not been introduced yet. With the help of a unique dataset of projects active from the beginning of the platform *HitHit* in 2012 till the end of June 2018, we will study the factors influencing the funding success of projects across all categories. These factors are analysed by the application of a logit and OLS model as is standard for similar studies.

Our results confirm most of the findings of global crowdfunding research. Unfortunately, as we did not have access to as much data as other researchers, the results of some of the hypotheses on the global scale remain unknown to the case of the Czech Republic.

This bachelor thesis is structured as follows. Chapters 2 and 3 focus on the existing literature about crowdfunding. While in Chapter 2 we focus on the theoretical concepts of crowdfunding and useful definition for a complete understanding of the whole theory of crowdfunding, in Chapter 3 we will focus more on the specific factors that affect successful funding of a project.

In Chapter 4, we will begin with the description of the dataset, the moti-

¹www.kickstarter.com

²www.hithit.com

³www.cnb.cz

vation behind the selection of individual variables for the subsequent analysis. Additionally, descriptive statistics are provided.

In Chapter 5, the methodology behind our models is described as well as some drawbacks we might encounter and their potential treatment. What we learn in Chapter 5 will then be applied in Chapter 6. Moreover, the interpretation of results and some important remarks are also discussed.

Finally, the last chapter concludes our results and suggests possibilities for future research.

Chapter 2

Crowdfunding

The objective of this chapter is to introduce some theoretical concepts which form the basis for the following chapters and sections of this thesis. In the first section, the basic definitions will be presented, followed by the taxonomy of crowdfunding, as found in the literature. Subsequently, the phases of a crowdfunding campaign will be described. The last section will be focused on the economic context of crowdfunding with an emphasis on Europe.

2.1 Definitions

Crowdfunding is a concept originating from two other concepts, *microfinancing* and *crowdsourcing*. The main idea behind microfinancing is lending small amounts of money to borrowers who do not have access to formal banking. The term is mainly associated with providing financial support to people in countries suffering from poverty (Morduch 1999). Crowdsourcing, on the other hand, is when an organization or an individual, work with crowds to obtain material, ideas or to complete various tasks in the form of an open call on the Internet (Kleemann et al. 2008). Instead of trying to attract a smaller number of specialists, whether they are investors or craftsmen, crowdsourcing as well as crowdfunding entrepreneurs aim to attract a broader audience.

To clarify, the terms *project creator*, *project funder*, and *crowdfunding platform* need to be defined. A *project creator* is an organization or an individual who creates and publishes a particular project on a *crowdfunding platform* (such as Kickstarter, IndieGoGo, Hithit, etc.) and therefore manages it during the project process, i.e., when the project is available to be funded online. On the other hand, a *project funder* is usually a person who, for various motives,

makes his funds available to support the particular project. For the most part, fundraisers are small to medium entrepreneurs or start-ups who do not have enough capital to finance their projects by themselves. It can be merely any legal or natural person with an attractive idea which they are able to elaborate and present to potential funders appealingly.

Crowdfunding platforms create an essential space where project creators who lack capital and project funders who are willing to invest can connect and communicate. The platforms set an array of rules and conditions for both parties involved in the project funding. Each platform can charge a different commission for the services it provides to the consumer, i.e., the project creator. The commissions are usually around 10%, the Czech platform Hithit charges a maximum commission of 9% for projects up to CZK 200,000 and a negotiable commission for higher priced projects. The creator only pays the specified percentage to the platform in case that the project is *successfully funded*.¹

When talking about a project being *successfully funded*, it is meant that the project reached the asking amount of money in the time frame it had been either given by the platform or the one that the project creator had chosen having been given a choice. Some platforms offer different time frames for different goal amounts. Usually, it is around 40 days, but rarely less than 15 days.

2.2 Taxonomy

Crowdfunding projects have various motives and objectives. They also vary in the nature and intensity of funding (Mollick 2014). According to De Buysere et al. (2012), there are four types of crowdfunding, donation-based, reward-based, lending-based, and equity-based. This thesis aims its attention at reward-based crowdfunding. However, a brief description of all types will be given. Massolution (2015) divides those four types into two more general categories, financial (lending-based, equity-based) and non-financial (donation-based, reward-based).

2.2.1 Financial Crowdfunding

Project creators in financial crowdfunding offer the contributors a possibility of a financial return at some point in the future. The two primary models

¹<https://www.hithit.com/cs/article/faq>

serve as alternatives to traditional funding: lending- and equity-based model of crowdfunding.

Lending-based crowdfunding is very specific. This type is relatively similar to loaning money in a bank. In this case, however, the customer is loaning money from a small group of creditors in place of loaning from one large creditor, the bank. The loans are, therefore, more accessible, and the interest rates tend to be lower. This is not the case of project funding, considering that funders lend their funds with the intention of their appreciation in the future. We can further divide lending-based crowdfunding into social lending and peer-to-peer (P2P) lending. Social lending operates predominantly in developing countries, where organizations can loan money without the commitment to pay interest. The second case, P2P lending, is based on the fact that the debtor and the creditor do not know each other. The primary motivation of the debtors is loaning with a lower interest than they would be offered in a bank, whereas the motivation of the creditors is a higher yield (De Buysere et al. 2012). The annual interest rates at the most famous Czech P2P lending platform Zonky.cz² start at 2.99% p.a.. Other worldwide popular P2P lending platforms include; Prosper, Funding Circle, or Zopa.

The equity-based model is the most recent one out of the four. The funder, more specifically, the investor in this case, has an opportunity to attain an equity instrument that provides him with a share of future earnings or even ownership itself (Massolution 2015). In this case, the investor's financial return entirely depends on the success of the business he invests in. For many people, equity-based crowdfunding represents an accessible form of investing without being a professional investor. On the one hand, the return from investment can be considerably higher than from investing in other types of crowdfunding. On the other hand, these investments carry by chance a higher risk. Moreover, for equity-based crowdfunding the legislation is likely to be more complicated in the majority of countries, in some, the scope of equity-based crowdfunding can be limited entirely (Belleflamme et al. 2010).

2.2.2 Non-financial Crowdfunding

The roots of crowdfunding come from charitable causes. The giving party would fund projects without any expectation of a benefit from the requesting party (Massolution 2015). For the most part, non-profit organizations who act as the

²zonky.cz

project creators would seek money from funders with similar interests who are interested in donating money for charitable purposes. GoFundMe offers two different methods of a donation-based campaign. First, a personal campaign where the funds collected are immediately available to the funder with the possibility of withdrawal at any time. The funder however, still needs to meet the conditions of the platform. Second, a certified charity campaign where the funds instead of reaching the campaign creator, go directly to the charity. The charity needs to be registered in the UK, the US, Australia, or Canada (GoFundMe 2018). The most popular donation-based crowdfunding platform worldwide is GoFundMe³ with over \$5 billion raised. The most well-known donation-based platform in the Czech Republic is Penězdroj.cz⁴ with the total amount of CZK 1.54 million raised.

The last of the four models is reward-based crowdfunding. In the campaigns, the funders are offered rewards or benefits for their contribution. The rewards may vary. Frequently, the project creators present a number of them on the project webpage. As an example, when the goal of the campaign is to finance the production of a CD, the rewards will surely involve a pre-order of the CD. In addition, for a higher price, it may also include tickets to a concert or a meet&greet of that particular music group. Typically, the more significant is the contribution, the better the prize.

With reward-based crowdfunding, the *all-or-nothing* (AON) financial scheme is most common. In this context, the project creator only receives the collected amount of money in case his project reaches the asked amount. In the other case, the money returns to the funder. For the other not so common *keep-it-all* (KIA) financial scheme, the fundraiser can keep the money collected despite not having reached the amount asked in the beginning.

In the case of all-or-nothing model, after a successful end of a campaign, the project gains independence from both the platform and the investors. The only obligation is then for the creator to bring the rewards to completion by delivering them to the funders in time.

2.3 Economic context

Crowdfunding has become a popular method of financing new businesses and creative projects and therefore, also an impactful part of the economy. This

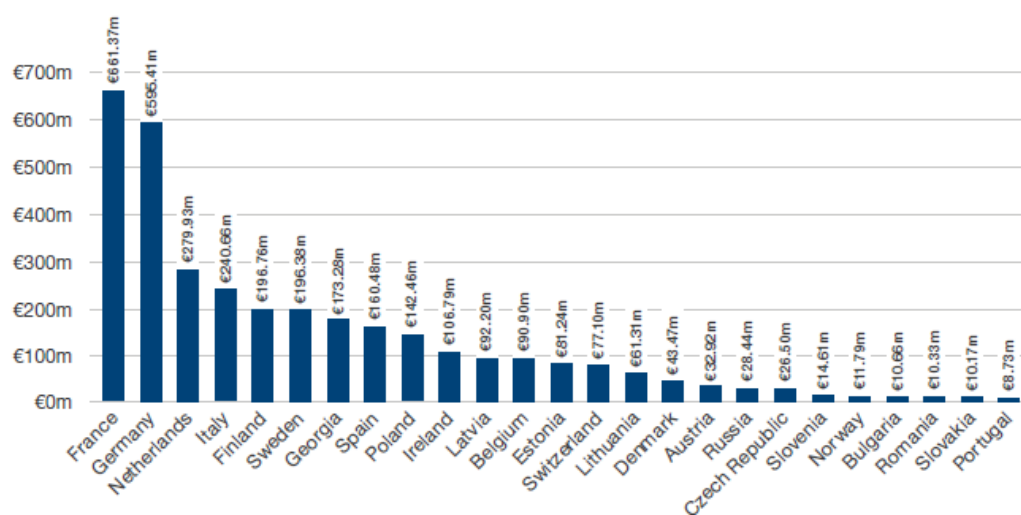
³www.gofundme.com

⁴penezdroj.cz

section will be dedicated to the economic context of crowdfunding overall and the position that the Czech Republic has compared to other economies of the European Union.

Alternative finance industry is continuously growing in Europe, the platforms are subject to regulations, and they keep expanding both nationally and internationally. Researchers from Cambridge University published the 4th European Alternative Finance Benchmarking Report (Ziegler et al. 2019), which will be used as core literature in this section. The study collected data from 269 platforms actively operating in 2017. These platforms come from across 45 countries in Europe. Nevertheless, not all of these are crowdfunding platforms. The largest market share belongs to P2P Consumer lending. Reward-based crowdfunding, on the other hand, accounted for 5% of the market share of European Alternative Finance. In Figure 2.1 we see where the Czech Republic stands in terms of alternative financing across other countries with €26.5m. In the Czech Republic, the reward-based crowdfunding remains the most common model (*Crowdfunding by Country* 2017).

Figure 2.1: Online Alternative Finance Volume by Country in 2017 (€millions)



Source: Ziegler et al. (2019)

The year 2017 has been the first one where reward-based crowdfunding's volume decreased. While in 2016 it was €191 million, the next year it was €159 million. This phenomenon could be due to the increased interest in investment model platforms in equity-based crowdfunding or donation-based platforms, where the projects are more likely to be successful because of the KIA policy.

The feature that makes the reward-based model more intriguing to study is indeed the fact that the campaigns can fail to be successful by a minimal volume of capital.

In 2017, the most successful sectors of reward-based crowdfunding in Europe were: 'Cultural and Creative Industries' with 35%, including sectors such as Music, Design or Arts, Secondly, 'Media and Publishing' with 12% and thirdly, 'Charity and Philanthropy' with 8%. These percentages show proportions of successfully funded projects on all the surveyed platforms.

In Eastern Europe, including Hungary, Poland, Slovakia, and the Czech Republic, the alternative finance market had increased by 153% from €70 millions in 2016 to €179 million in 2017. Conversely, the Czech Republic's market stagnated going from €31 million in 2016 to €27 million in 2017. It could be partially explained by the absence of a few platforms which did not participate in the survey in 2017, while they did in 2016 (Ziegler et al. 2019). The reward-based crowdfunding model is known to have raised only €250,000 in 2017 compared to €1.85 million in 2017.

While the market share of crowdfunding is still relatively small, it is to some extent regulated the same as other markets in the Czech Republic. Regarding the equity-based model, the intermediaries and consumer loan providers must obtain a license from the Czech National Bank. However, from reward-based crowdfunding, there has been an exemption, and it is not qualified as an Investment Fund; therefore, the reward-based model platforms do not require a license (Ellenoff 2017).

2.4 Crowdfunding Campaign Phases

To familiarize the reader with the process of crowdfunding a short summary of individual phases of a campaign needs be presented. This thesis is mainly focused on reward-based crowdfunding, which is why from now on the term reward-based crowdfunding will be referred to as crowdfunding only. It is necessary to remember that it is a tool of financing projects, especially creative ones such as publishing books, filming movies, the development of a software, or other creative inventions. In this model of crowdfunding, the funders never obtain any sort of share from the future earnings, hence it is not to be confused with the methods of financing of traditional entrepreneurial entities where there is profit expected.

Preparatory phase

It is crucial to decide whether the scope of the campaign is going to be oriented towards a local community or a larger crowd, either nationally or multi-nationally. Besides that, each campaign has different goals. According to Mollick (2014), there are two main types of goals. The first one usually concerns one-time projects where the creator is asking for a relatively small capital, often under \$1000. In this particular case, the capital is often provided by friends and family. However, raising initial capital to fund new ventures of an entrepreneur is becoming more and more popular. For example, forty-five out of the fifty highest funded projects were of the second type and even successfully continued as entrepreneurial firms. Besides these, creators can also use crowdfunding as a way of learning if there is sufficient demand for the presented project or product as was the case of one Canadian smartwatch start-up which was then able to secure more funding via a more traditional source (Dingman 2013). Therefore, their primal goal was not funding through the crowdfunding platforms.

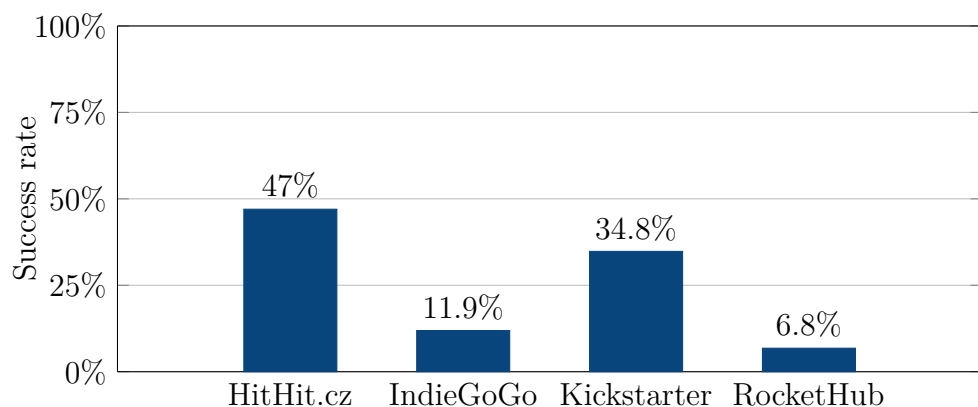
From thereon, a project creator can start deciding which crowdfunding platform would be the most suitable for his specific project.

The project creator has an option to either join one of the crowdfunding platforms already available on the Internet, or he can choose to set up his own web page which would specialize in an area closer to the objective of his campaign. To be able to do that, however, one must already have substantial capital. Hence, this case would instead apply to projects with a considerable and already existing community of fans. Nowadays, most of the project funders can choose between several platforms considering their various specializations, visual aspects of their website, public awareness, or the fees of using their services (Steinberg 2012).

One of the relevant aspects could also be the success rate of a platform. In Figure 2.2, the success of a platform is depicted by the ratio of successfully funded projects to the total number of projects created on a platform. Using this measure, the Czech platform Hithit.cz has very favourable results compared to its foreign competition. It must be however mentioned that worldwide crowdfunding platforms operate with a significantly larger crowd than the ones that operate locally, such as HitHit.

Once the platform is chosen, the creator can step to preparing the campaign itself. Although the campaigns differ in many ways, their tools and objectives are almost the same. Thus, the upcoming procedure should be the same for

Figure 2.2: Success rates of crowdfunding platforms



Source: www.statista.com

most of the projects.

The first important thing is to create a prominent project description. This serves as the main way to present the project to potential funders. Second, the creator should decide if he wants to upload a video that would complement the project description.

Yet another significant part of the preparatory phase is choosing the appropriate amount of rewards and their value. Sometimes, a funder may be interested in other products than the one which is fundamental to the project. Thus it is crucial to focus on the complementary goods just as much as on the main product.

After having collected the material needed to set up a campaign, the creator needs to think about how much money he needs to raise. The amounts can vary from as little as 50\$ to as much as thousands of dollars.

The next step is to set the individual rewards. Usually, the projects offer a pre-selling of the product and some complementary products as well. Also, one of the popular rewards is a simple postcard, for the smallest of contributions, or a simple thank-you note after the end of the campaign. It is essential to find the right balance between tangible and intangible rewards that are creative and reasonably priced.

Finally, the last move is to set the duration of a campaign. In many cases, this is decided by the crowdfunding platform. It may depend on the goal amount of the project, or the platform may have a fixed time duration for all of the campaigns. The Czech platform *HitHit* gives a choice of 30 or 45 days.

Once everything is set, the project creator uploads the project onto the platform which then assesses it and publishes it (*HitHit.cz - FAQ 2019*). Al-

ternatively, they may return it to the creator with some recommendations for improvement.

Campaign in progress

After the project is published, the creator is advised to remain active during the campaign. One of the steps, he can take is attracting potential funders via social networks. Additionally, the creator tries to be as active as possible to the extent that the platform allows.

Some platforms allow for updating the campaign, such that the backers know that the project is on track. Creators are sometimes allowed to add new rewards, answer the backers' questions, and publish these answers on the campaign web-page. The response rate should be swift due to the limiting duration of the campaign (Ennico 2016).

Final phase

Once the period allowed for collecting funds has ended and the campaign has been successful, the creator thanks all the contributors and waits a certain time, set by individual platforms, before being able to receive the collected funds. He pays a commission that has been agreed upon and receives a list of contributors and their respective rewards. In the other case, when the campaign has not been successful, the collected money would return to the funders. It is still crucial to remain in contact with the backers so that they do not lose confidence in the project. In case the creator wants to return to crowdfund another project or an improved version of his initial one, having a stable community base is valuable.

Chapter 3

Literature Review

This chapter is a review of already existing literature related to the purpose of this thesis. There have been many of studies published mostly focusing on the global crowdfunding platforms, or only specific categories of projects. These studies will be mentioned and briefly described in this chapter. The first section will take into account factors that could be affected by the project creator. The second section will then discuss studies and papers, mentioning factors that the creator may not be able to affect.

3.1 Factors influenceable by the creator

This section will cover only factors that the project creator can influence before or during the campaign. It will include mentions of studies focusing on the characteristics of a description, categories of projects, the presence of a video, and other factors, or models that are able to predict the success of a campaign with a given certainty.

According to Mitra and Gilbert (2014), language and the content of the description are of great importance in connection with the success of a project. The creator should avoid using negatively sounding sentences such as; 'Only a dollar less, and we will lose everything.' On the other hand, the creators should focus on offering empathy and gratitude to the ones supporting them. One also needs to be careful of spelling errors; those unsurprisingly affect the success of a campaign negatively (Mollick 2014). The use of words such as '*exciting*' or '*happy*' or even the sense of inclusion when using the word '*together*' or the pronoun '*we*', make a campaign more attractive. Typically, the use of these words is highly associated with projects of female founders. As Gorbatai and

Nelson (2015) suggest in their study, women, in general, tend to raise more money in crowdfunding projects than men.

Moreover, the length of description is also an aspect that is connected to the success of a project (Bi et al. 2017). It was found that longer descriptions tend to have a positive effect on campaigns. The study uses a dataset from a Hongkongese crowdfunding platform *zhongchou*¹ which has become one of the largest crowdfunding platforms in China. The team conducted a hierarchical regression analysis as they tried to find the most significant influence factors. One of the strongest signals of project quality was, in fact, the introduction word count.

The study of Greenberg et al. (2013) suggests a model that can predict the success of a campaign with a 68% accuracy based on the attributes it is given before the launch and with 89% accuracy after 15% of the time of the campaign duration has passed. The model construction was rather complicated. It built several machine learning algorithms that work together. In their future works, they would like to build a feedback tool that gives advice during the creation of a campaign, such as: "We noticed you had not uploaded a video. Projects with videos tend to be 10% more successful."

The presence of a video in a campaign has not been the subject of only one study (Mollick 2014; Bi et al. 2017; Belleflamme et al. 2014). It has been shown that on the studied platforms, in most cases, global platforms such as Kickstarter and IndieGoGo, the presence of a video has a positive effect on the success of the project campaign. It is vital to be brief, but striking, in order to get the point of the campaign across (Steinberg 2012). The majority of studies focusing on reward-based crowdfunding conduct analyses using logistic regression controlling for mostly the same factors depending on the access to the data. A variable that none of them is missing is, however, the goal amount of a project.

Stegmaier (2015) suggests that the amount should be the smallest possible, sufficient for the implementation of the project only. Projects with goal amounts set inadequately high tend to be less successful. In other studies (Mollick 2014; Chan et al. 2018), increasing goal amount is also negatively associated with the success of a project.

Another factor, where we would expect a significant effect on success is the variety of offered rewards. Nevertheless, analysing it is nearly impossible because of the great variety. For example, the pre-selling of a product is very

¹www.zhongchou.cn

popular and also appealing (Burtch et al. 2013). However, it is not the only important reward the project should offer. Here, price discrimination between regular customers and the crowdfunders often presents itself. It is a smart way to attract funders. However, when the distortion in the price discrimination is overdone, the project may start to lose its appeal (Belleflamme et al. 2014).

Even though it would seem logical that the longer the duration, the higher chance of collecting all the necessary funds, it has been shown that shorter durations give the creator a sensation of urgency which will motivate him to find resources faster (Stegmaier 2015). The optimal duration for entrepreneurs starting with crowdfunding is said to be around 25 days.

Some studies concentrate only on factors influenceable before the publication of a campaign on the platform. Necessarily, the factors that can be affected after the launch are just as important to study as the ones in the preparatory phase. In their study, Koning and Model (2013) divide new projects into categories by contributions: small contribution, large contribution, no contribution. Then they suggest that higher initial amounts funded to the project are more beneficial than smaller ones. They also propose that no initial funding is better than having the project funded by small amounts in the beginning. They do, however, take data from a donation-based crowdfunding platform. But it could be easily applicable to other types of crowdfunding, including reward-based. Some creators may choose to be the first ones to fund their projects with relatively high contributions, either purchasing the rewards or contributing to an intangible reward, in order to attract other funders supposing that when there is a significant part already funded, it may attract more investors. A strongly related factor could also be the size of the creator's Facebook community.

The success of a campaign is profoundly affected by one's social network activity (Mollick 2014). One specific example may be that a high number of Facebook friends or fans have a positive effect on the outcome of the campaign. According to Young (2012) only 14% of customers confide in advertisement found on-line or on TV, whereas 78% have more confidence in personal recommendations. Having a Facebook friend sharing a crowdfunding project gets more of a person's attention than seeing an ad posted by Facebook.

During a campaign, the creator usually has access to communication with the customer. It is essential to post updates or answer frequently asked questions for the project to be more successful (Ennico 2016).

3.2 Other influential factors

While there are factors that can be influenced in either one of the stages of crowdfunding, there are also elements whose affecting is not in the power of the project creator. These factors will be described in the following section of Chapter 3.

First of these factors is extensively described in the work of Agrawal et al. (2011) '*The geography of crowdfunding*'. It suggests that the geographical proximity of the funder and the entrepreneur is of importance. Furthermore, projects that are based in larger cities or more densely populated areas also tend to be successful. In their later study, Agrawal et al. (2015) suggest that local funders, those who live in the proximity of the project creator, tend to invest more at the beginning of the project duration, whereas the tendency of distant funders to invests increases with the amount collected. This is explained by the fact that at the beginning of a project, the majority of funds come from friends and family.

Chapter 4

Data

This chapter focuses on data collection and selection for later analysis. First, a number of dependent and independent variables used in the models are presented. Second, several hypotheses are introduced along with the variables needed for the estimation. Finally, the descriptive statistics are provided to familiarize the reader with the dataset properly.

The Czech crowdfunding platform HitHit provided me with a dataset of cross-sectional data. Additional data needed for the analysis were scraped from the webpage of each crowdfunding campaign analysing the Document Object Model (DOM) using a PHP script.

4.1 Data selection

From the comprehensive dataset gathered by HitHit, not all of the 2044 observations were selected for my analysis. The final dataset includes both successful and unsuccessful project campaigns with their specific information. These include the duration, asked goal amount, the collected amount of money, number of funders and rewards, average contribution per project, division into categories, information about the presence of a video, and some feature a location where the project founder resides.

Additional data extracted from the websites are the length of the project description, the number of updates during the campaign, and also the number of questions answered and published on a project's page.

In order to dispose of possible outliers, I only used the data for the number of backers being higher than 1. In cases where the number of backers is either 0 or 1, we can suppose that the project founder decided to abandon his project

right in the beginning. Since I choose to study the determinants before and during the campaign, 61 projects are considered unsuitable for my analysis and were left out.

During the data collection from the website, I encountered that some of the projects do not have any characters in their descriptions. After checking them individually, I learned that instead of a text description, the project founder had uploaded a picture with the description in it. This, however, makes it too complicated to count the actual number of characters. Hence, 112 projects were also disregarded in the final dataset. Consequently, 1871 observations remained.

4.2 Variables and Hypotheses

In this section, I present the dependent and independent variables used in the analysis and the motivation behind choosing them. The goal of this thesis is to analyse the cross-sectional data of one of the most popular crowdfunding platforms in the Czech Republic based on the success of individual projects with their characteristic variables. Moreover, hypotheses connected to the presented independent variables are introduced as well and will be tested later on.

Dependent variables

Table 4.1: Dependent variables

Variable	Description
<i>funded</i>	Information whether the project succeeded or not
<i>success</i>	Percentage of collected capital (can exceed 100%)
<i>average_contribution</i>	Average contribution of a project per one backer

In order to differentiate between the logistic and OLS regression, three dependent variables are chosen for the analysis. The dependent variable *funded* determines whether the project reached its goal amount in the original time frame. This binary variable will be used for logistic regression. Another variable, *success*, which states the percentage of obtained funds during a campaign, will be presented in the OLS analysis. In our case, the percentage can exceed 100; nevertheless, it can never be negative. Lastly, the variable *average_contribution* presents the average contribution of one backer for each

project and will be only mentioned marginally in the analysis. The summary of used dependent variables can be seen in Table 4.1.

Independent variables

The independent variables include factors that can be affected by the project creator and are easy to measure or quantify for our purpose. In Table 4.2, basic information about each independent variable is summarized. Some variables need more clarification, which will be given with the corresponding hypotheses.

Table 4.2: Independent variables

	Variable	Description
	<i>goal_amount</i>	The amount of asked capital (in thousands of CZK)
	<i>rewards</i>	Number of rewards in each project
	<i>updates</i>	Number of updates during the campaign
	<i>qa</i>	Number of answered questions during the campaign
	<i>word_count</i>	Number of words in the description
	<i>backers</i>	Number of backers of the project
	<i>has_video</i>	Dummy variable for a video in the project
	<i>short_campaign</i>	Dummy variable for a campaign duration of 30 days
	<i>large_city</i>	Dummy variable for location within a large city
Category	<i>technology</i>	Dummy variable for technology
	<i>arts</i>	Dummy variable for arts
	<i>education</i>	Dummy variable for education
	<i>entertainment</i>	Dummy variable for entertainment
	<i>community</i>	Dummy variable for community
	<i>miscellaneous</i>	Dummy variable for other categories

Hypothesis 1

Increasing the amount of money asked for the project has a negative impact on the success of project funding.

Goal_amount is a variable that defines the amount of money a project creator has asked for at the beginning of a campaign. It cannot be changed during nor after the end of the campaign. This variable was chosen because projects with very high or even unrealistic goals could be assumed not to attract enough backers to fund the campaign. Studies mentioned in Chapter 3 had also studied this effect, as well as effects defined in the following hypotheses. The

results that hold in other countries, and globally, will be studied for the case of the Czech Republic.

Hypothesis 2

The number of rewards has a positive impact on the success of the project.

In this case, the hypothesis will be tested with a focus on the variable *rewards*. One would assume that the more varied the offer, the higher is the chance of a backer to choose his ideal reward. This is why I assume a positive impact.

Hypothesis 3

Updating the project page has a positive impact on the success of project funding.

Hypothesis 4

Answering questions relevant to the project has a positive impact on the success of the project.

Being more engaged in the whole process of a campaign seems beneficial for the success of a project. Therefore, two hypotheses to be tested is stated. In a purely logical sense, updating the backers about how the project is doing, what are the plans for the future, or informing about the approximate time of delivery of the rewards, gives the backer a sense of inclusion in the process. This should make the campaign more attractive to other backers as well. The same thing applies to the number of answered questions on the web page of a campaign. Variable *updates* specifying the number of updates during a campaign and variable *qa* defining the number of answered questions on the webpage of a campaign will be used.

Hypothesis 5

The length of a project description has a positive impact on the success of the project.

Considering that the backer receives most of the information about the project from the product or project description, longer descriptions, or potentially more informative descriptions, tend to have a positive impact on the success of project funding (Bi et al. 2017), as described in Chapter 3. I will test this hypothesis for the case of the Czech Republic in like manner with the help of the independent variable *word_count*.

Hypothesis 6

Having a project listed in the category technology has a positive impact on successful project funding.

This hypothesis includes the variables describing categories. Since initially, a dataset with 16 different categories was provided, I decided to reduce the number of categories in order to make the results of the regressions clearer. The dummy variable for category *technology* includes project classified in original categories *technology* and *impact hub*, *art* includes project from art, design, or fashion, *education* has project from original category *education* as well as writing, and 'letniskola' which is a particular category including only a small number of projects connected to an educational film camp in the Czech Republic. The dummy variable *entertainment* also includes 'letniskola', movies, theatre, games, or sports-related projects. The variable *community* remained the same, and *miscellaneous* includes projects which I was unable to classify into one of the previous categories, it includes food-related projects and projects sponsored by Vodafone. Each project can be classified into one or two categories; nevertheless, each project has at least one dummy variable for category equal to one. The hypothesis was chosen due to the society's constant thrive for innovation. Intuitively, we would assume that projects with a technological theme would be more attractive to the backers.

Hypothesis 7

The presence of a video has a positive effect on successful project funding.

Hypothesis 8

The location of a founder within a large city has a positive impact on the success of the project.

The two previous hypotheses stem from the previously studied literature and include the rest of the independent variables used for the analysis of determinants of success of reward-based crowdfunding projects (Agrawal et al. 2011; Mollick 2014). Used variables are the presence of a video, *has_video*, and the location of the project creator, using a dummy variable *large_city*. Since not all projects had the data on location available, I decided to create a new variable that would inform only about the location being in one of the largest cities. Because the purpose of our study is the Czech Republic, a city is considered large when it is one of the three largest cities in the Czech Republic, i.e., Prague, Brno, and Ostrava.

Hypothesis 9

Shorter campaign durations have a positive impact on successful project funding.

Referring to Stegmaier (2015), I will test this hypothesis in order to learn whether campaigns in the Czech Republic with the shorter option of duration tend to be more successful than those with other durations. In the dataset, the two main possibilities are 30 or 45 days. However, some irregularities also occur, where the value was neither 30 nor 45, which is why I decided to define the dummy variable *short_campaign* as being equal to 1 when the duration is 30 days, and 0 otherwise.

4.3 Descriptive statistics

Table 4.3 shows the descriptive statistics of the three considered dependent variables.

Table 4.3: Descriptive statistics/dependent variables

Statistic	Mean	St. Dev.	Min	Max
funded	0.510	0.500	0	1
success	0.673	0.621	0.001	7.185
average_contribution	856.764	856.270	4	22,367

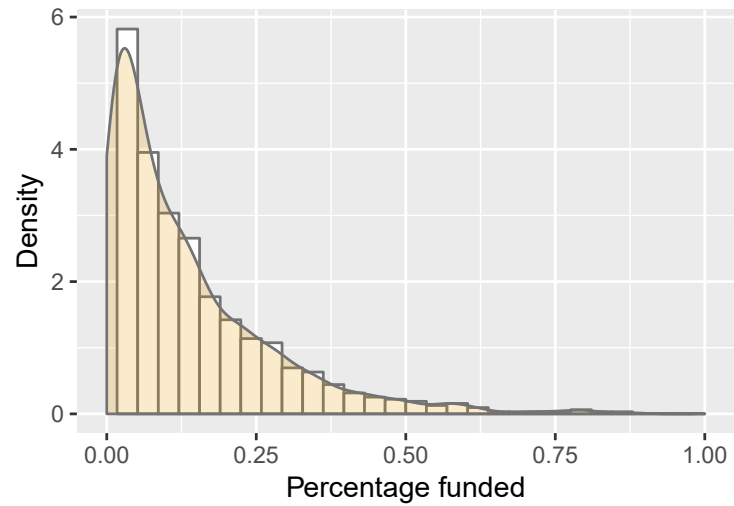
N=1871

Even though the ratio of successfully and unsuccessfully completed projects is somewhat balanced, the interesting thing is that the distribution of successful and unsuccessful projects is in a way similar. Both can be seen in Figures 4.1 and 4.2. Hence, successful projects tend to succeed by small margins, whereas unsuccessful projects tend to fail by relatively large margins. The average funding level for unsuccessful projects is 13%; for successful, it is 119%. Only 34 projects from our dataset managed to collect more than 200% of the goal amount.

The following table 4.4 represents all the 15 independent variables considered in the models and their respective descriptive statistics.

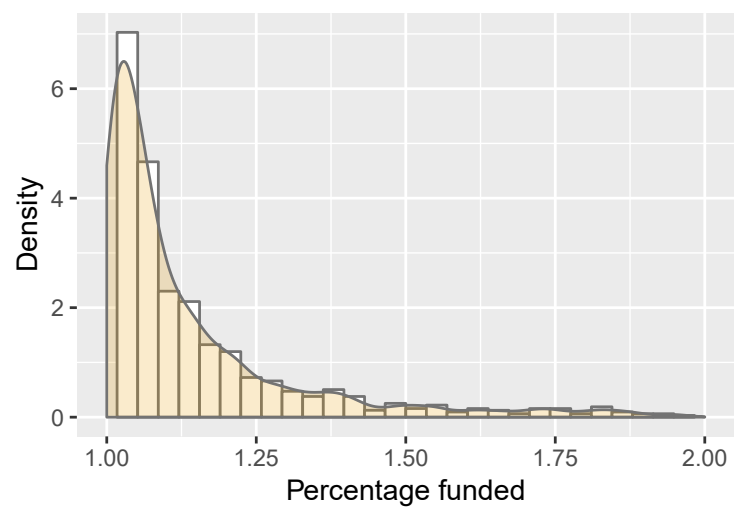
As can be seen from the table, a large proportion of variables is binary. Consequently, the mean values measure the proportion of a variable in the dataset. Almost 37% of project founders are located within a large city. Nevertheless, this number is only illustrative because, for some of the projects, the

Figure 4.1: Funding levels of unsuccessful projects



Source: author's computations

Figure 4.2: Funding levels of successful projects



Source: author's computations

Table 4.4: Descriptive statistics/independent variables

Statistic	Mean	St. Dev.	Min	Max
goal_amount	119.869	144.607	2	2,000
rewards	13.846	7.609	2	82
updates	2.835	3.929	0	32
qa	0.890	1.876	0	16
word_count	514.117	236.300	32	1,478
backers	96.373	185.085	2	3,634
<i>categorical variables</i>				
technology	0.060	0.238	0	1
arts	0.257	0.437	0	1
education	0.284	0.451	0	1
entertainment	0.463	0.499	0	1
community	0.068	0.253	0	1
miscellaneous	0.091	0.288	0	1
<i>other dummy variables</i>				
has_video	0.820	0.384	0	1
short_campaign	0.130	0.337	0	1
large_city	0.369	0.483	0	1

N=1871

information on the location was missing. For dummy variables *short_campaign* and *has_video* the summary statistics are very straightforward. While many project creators decide to include a video in their campaign, only a small proportion of them, 13%, decides for the shorter duration of a campaign. It must be pointed out that projects with shorter duration campaigns do not necessarily have lower goal amounts.

The majority of projects are included in only one of the categories; however, 12.6% of the projects fall into two categories. The most substantial proportion of projects fall into the entertainment category, which includes products such as movies, games, or sports-related projects, as described in the previous section.

The variable *goal_amount* is measured in thousands of Czech crowns. Since the lowest asked amount is 1,700 CZK, scaling the variable made perfect sense. Only 11 projects from our dataset set their goal amounts higher or equal to 1 million Czech crowns. On average, a project founder asks for an amount of almost 120,000 CZK.

Moreover, he offers on average 14 distinct rewards. The minimum number of rewards offered in a project is 2, the maximum is 82. Some projects do not post any updates on their webpage, an average project has around 3 posts. The number of updates can reach up to 32. Answering questions and posting them

on the project page is not very common. Out of 1871 projects, 1329 had not answered any questions. For those, who did, the average number of answered questions is 3.

The longest description recorded of a project was 1,478 words, in other words more than five pages. 88% of projects have a description longer than one page (260 word per page). The shortest description was 32 words while the average description length is 514 words, around 2 pages. The variable *word_count* was obtained by dividing the number of characters in a description by the average number of characters in Czech words.

Chapter 5

Methodology

This chapter is focused on the description of the models and variables used for the estimation. Two models are considered. First, the logit model to study the factors affecting the success of a crowdfunding campaign in the Czech Republic. Most authors use logistic regression as a tool to determine the influential factors. Second, the OLS model, used as a complement to the logit model, will study the factors influencing the success of a project on the whole scale. The success variable used in the OLS model is measured in percentage of the goal amount; thus, it can go over 100%, reaching up to 719% of the goal amount for the 'most successful' project. The theoretical background of this chapter is based on Wooldridge (2015).

5.1 Logit model

The dependent variable *funded* recognizes only cases when the project was successfully funded (=1), thus reached the initially asked goal amount, or it was not (=0). One approach to model this situation would be the linear probability model (LPM), which uses the methods of classical OLS. However, LPM has some limitations. The two most severe ones are: the fitted probabilities may reside outside the interval of (0,1), and the heteroskedasticity of the error term might be present. To avoid these limitations, the logit model was selected.

The logit model is a *binary response model* where we are mostly interested in the response probability

$$P(y = 1|\mathbf{x}) = P(y = 1|x_1, x_2, \dots, x_n)$$

where the full set of explanatory variables is denoted by \mathbf{x} . To avoid the

limitations posed by LPM, we consider a function G which takes on values strictly between 0 and 1: $0 < G(z) < 1, \forall z \in \mathbb{R}$.

The function G , in its general binary response model form, is defined as follows:

$$P(y = 1|\mathbf{x}) = G(\beta_0 + \beta_1x_1 + \dots + \beta_kx_k) = G(\beta_0 + \mathbf{x}\boldsymbol{\beta}). \quad (5.1)$$

This form ensures that the values of the estimated response probabilities lie strictly between zero and one. For the logit model, G is the logistic function:

$$G(z) = \frac{\exp(z)}{1 + \exp(z)} = \Lambda(z). \quad (5.2)$$

The cumulative distribution function for a standard logistic random variable is the same. Moreover, the logistic cumulative distribution function has a very similar shape to the standard normal one. The function G is an increasing function. It increases most quickly at z being equal to zero. As $z \rightarrow -\infty, G(z) \rightarrow 0$, and as $z \rightarrow \infty, G(z) \rightarrow 1$.

Using a latent variable model, we will derive the logit model. For this, let y^* be an unobserved, latent, variable such that

$$y^* = \beta_0 + \mathbf{x}\boldsymbol{\beta} + e, \quad y = 1[y^* > 0]. \quad (5.3)$$

The notation $1[\cdot]$ defines an *indicator function*; it is equal to one when the event in the brackets is true, and zero when it is not. In other words, $y = 1$ if $y^* > 0$, and $y = 0$ if $y^* \leq 0$. We assume that e has the standard logistic distribution and that it is also independent of \mathbf{x} . From the assumptions and Equation 5.3, we derive the response probability for y :

$$\begin{aligned} P(y = 1|\mathbf{x}) &= P(y^* > 0|\mathbf{x}) = P[e > -(\beta_0 + \mathbf{x}\boldsymbol{\beta})|\mathbf{x}] \\ &= 1 - G[-(\beta_0 + \mathbf{x}\boldsymbol{\beta})] = G(\beta_0 + \mathbf{x}\boldsymbol{\beta}). \end{aligned}$$

We can notice that we yield the same result as in Equation 5.1. Using it and plugging it into Equation 5.2, we get:

$$P(y = 1|\mathbf{x}) = G(\beta_0 + \mathbf{x}\boldsymbol{\beta}) = \frac{\exp(\beta_0 + \mathbf{x}\boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}\boldsymbol{\beta})}. \quad (5.4)$$

The interpretation of beta coefficients is not as straightforward in the logit model as it is in the linear probability model due to the nonlinear nature of

$G(\cdot)$. For our purpose, we want to estimate the effects of x_j on the probability of success. In order to find the partial effect of the variables on the response probability, we need to take the partial derivatives. We find the partial effect on $p(\mathbf{x}) = P(y=1|\mathbf{x})$ as follows:

$$\frac{\partial P(\mathbf{x})}{\partial x_j} = g(\beta_0 + \mathbf{x}\boldsymbol{\beta})\beta_j, \quad \text{where } g(z) = \frac{dG}{dz}(z). \quad (5.5)$$

The function g is the probability density function. And since $G(\cdot)$ is a strictly increasing function, $g(z) > 0 \quad \forall z$. Thus, the partial effect of x_j will always have the same sign as β_j for all j . Therefore, it is easily determined whether an explanatory variable has a positive or negative effect. However, estimating its magnitude is more complicated.

For the estimation of binary response models that are nonlinear, we use the Maximum Likelihood Estimation (MLE). The maximum likelihood estimators (MLEs) are under general conditions consistent, asymptotically normal, and asymptotically efficient. To obtain the MLE, we assume we have a random sample of size n that is large enough, and the density of y given x_i .

$$f(y|\mathbf{x}_i; \boldsymbol{\beta}) = [G(\mathbf{x}_i\boldsymbol{\beta})]^y [1 - G(\mathbf{x}_i\boldsymbol{\beta})]^{(1-y)}, \quad y = 0, 1$$

To derive the log-likelihood function, we take the logarithm of the previous equation.

$$\ell_i(\boldsymbol{\beta}) = y_i \log[G(\mathbf{x}_i\boldsymbol{\beta})] + (1 - y_i) \log[1 - G(\mathbf{x}_i\boldsymbol{\beta})] \quad (5.6)$$

As our function $G(\cdot)$ lies in the interval $(0,1)$, $\ell_i(\boldsymbol{\beta})$ is well defined for all β values. To obtain the log-likelihood function for a sample size of n , we sum the Equation 5.6 across all observations such that $\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\beta})$. The MLE of $\boldsymbol{\beta}$ denoted by $\hat{\boldsymbol{\beta}}$ is the *logit estimator*.

Since the logistic regression estimates are reported in the form of log-odds, it is complicated to interpret them. For interpretation purposes, we are offered some possibilities. The first of them is the interpretation with the help of the odds ratio. The idea is to exponentiate the coefficients and then interpret them.

Second, we can also use average partial effects (APE) or partial effects at average (PEA). The PEA serves for analysis of the 'average person' in the sample. Nevertheless, for binary variables, such as *has_video*, it does not make sense to say that the average project has 0.82 videos. In this case, the use of

APE is more reasonable as it averages the individual partial effect across the sample.

Unfortunately, the traditional R^2 cannot be used as a goodness-of-fit measure in the case of the logit model. For this purpose, the McFadden's pseudo R^2 is presented. Its value has similar properties as the traditional one; it can only obtain values on a scale from zero to one, and the higher the number, the better the fit of the model. It describes how much of the variation in data is explained. McFadden's pseudo R^2 is calculated as follows:

$$R^2 = 1 - \frac{L_{ur}}{L_r}$$

where L_{ur} stands for the log-likelihood function of the estimated unrestricted model and L_r stands for the log-likelihood function of the same but restricted model including the intercept only.

5.2 OLS model

In this section, the assumptions for the OLS model will be presented alongside with the compensations for found limitations. As this chapter will serve as theoretical background to our used method, several models in the following form will be introduced:

$$y_i = \beta_0 + \mathbf{X}\beta + u_i \tag{5.7}$$

where y_i is the dependent variable, in our case called *success*, which measures the percentage of collected money with respect to the goal amount of the project i . \mathbf{X} represents the vector of control variables, including the factors presented in the previous chapter, u_i is the error term that is assumed to be identically independently and normally distributed.

To be able to validate the cross-sectional model, the Multiple Linear Regression assumptions (MLR.1 to MLR.5) need to be satisfied. Those assumptions can be closely seen in Appendix A. As all parameters in our model are linear, the first assumption is satisfied. By including as many variables as possible, we control for no correlation of the error term with the explanatory variables. To satisfy the random sampling condition, we have a dataset from one of the

largest reward-based crowdfunding platforms in the Czech Republic, which is considered a relevant and random sample for the case of the Czech Republic.

Also, by including a square term in the regression, we pose a threat to the no perfect collinearity assumption. To control for that, we use centering of the given variable which relieves us of the problem. The application of the Variance Inflation Factor method was used to check for any linear relationships between variables. The table summarizing the results can be seen in Appendix A.

Finally, the assumption of homoscedasticity needs to be satisfied. By executing the Breusch-Pagan test, we suspect heteroscedasticity. This particular test detects whether the variance of u_i is dependent on the independent variables. To treat for heteroskedasticity, we have the option of using robust standard errors. By the rule of thumb, when having more than 100 observations, it is appropriate to use the robust standard errors and conclude that they are valid. Under these assumptions (MLR.1 - MLR.5) the OLS estimator is the best linear unbiased estimator (BLUE).

Chapter 6

Results

This chapter presents the results of run regressions and discusses the methodology that was described in Chapter 5 in connection to our particular case. The primary tool to analyse and estimate the data was the R software. We used two different approaches to regression analysis to estimate the impact of several project characteristics on the success of funding of a project in the Czech Republic. The two main studied aspects were the percentage of overall funding and the simple indication of whether the project was sufficiently funded to be executed. In each section, the individual models will be presented with their respective results and interpretation.

6.1 Regression results - the logit model

The motivation behind choosing the logit model for our estimation was that most authors of studies conducted about crowdfunding in all its forms use logistic regression. These authors include Mollick (2014); Belleflamme et al. (2010) whose works I mentioned in the literature review in Chapter 3, but also authors who do not focus on reward-based crowdfunding in general but aim their attention at more specific fields, such as music, film, etc., or even at different crowdfunding models.

In order to find the best fitting model, we ran a likelihood ratio (LR) test on two models, restricted and unrestricted, one with explanatory variables including the dummy variables and one without them. This test assesses the goodness-of-fit of two models based on the ratio of their likelihoods. The LR test is computed as follows: $LR = 2(\mathcal{L}_{ur} - \mathcal{L}_r)$. The \mathcal{L}_{ur} represents the log-likelihood of the unrestricted model and \mathcal{L}_r stands for the log-likelihood of the

restricted model. We multiply the difference by two to assure that the LR has an approximate chi-square distribution with q degrees of freedom, $LR \sim \chi_q^2$ (Wooldridge 2015). The results of the LR test can be seen in Table 6.1, where M1 is the restricted model and M2 the unrestricted model. The results from the two models are estimated in Table 6.2.

Table 6.1: Likelihood ratio test results

Likelihood-ratio test	$Chisq = 29.078$
(M1 nested in M2)	$\Pr(> Chisq) = 0.0006286$

After consideration of the factors influencing the success of a project in the relevant literature as well as the results from the previous test, the following model, given by Equation 6.1, has been proposed. The function G is a logistic function.

$$\begin{aligned}
 P(\text{funded} = 1|\mathbf{x}) = G(\beta_0 + \beta_1\text{has_video} + \beta_2\text{short_camp} \\
 + \beta_3\log(\text{goal_amount}) + \beta_4\text{rewards} + \beta_5\text{large_city} \\
 + \beta_6\text{technology} + \beta_7\text{arts} + \beta_8\text{education} \\
 + \beta_9\text{entertainment} + \beta_{10}\text{community} + \beta_{11}\text{miscellaneous} \\
 + \beta_{12}\text{updates} + \beta_{13}\text{qa} + \beta_{14}\text{word_count})
 \end{aligned}
 \tag{6.1}$$

In the following section, I study the estimated effects of independent variables. The estimation is executed in the same form as described in Chapter 5. The final model has 1,871 observations with a Pseudo R^2 equal to 0.171.

Table 6.2: Logistic regression

	<i>Dependent variable:</i>	
	funded	
	(1)	(2)
has_video		0.464*** (0.143)
short_camp		0.325** (0.156)
log(goal_amount)	-0.343*** (0.065)	-0.357*** (0.067)
rewards	0.074*** (0.009)	0.068*** (0.009)
large_city		0.271** (0.110)
technology		0.170 (0.240)
arts		-0.003 (0.140)
education		0.378** (0.161)
entertainment		0.303* (0.158)
community		0.211 (0.240)
miscellaneous		0.277 (0.206)
updates	0.292*** (0.022)	0.289*** (0.023)
qa	-0.054* (0.031)	-0.057* (0.031)
word_count	-0.001*** (0.0002)	-0.001*** (0.0002)
Constant	0.550* (0.289)	-0.142 (0.354)
Observations	1,871	1,871
Log Likelihood	-1,088.830	-1,074.291
Pseudo R ²	0.160	0.171
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

6.1.1 Interpretation of results

As was already described in Chapter 5, the interpretation of logistic regression is not as straightforward as the one of the linear probability model. In order to reach a state where we can easily interpret any results, we need to apply the average partial effects (APE). In table 6.3, we present the estimated coefficients of the selected model with their corresponding APE. As a robust check for the calculated APE values, we used LPM, the results of which are attached in Appendix A. For the LPM, we need to satisfy the same assumptions as in the OLS model. The first four assumptions are satisfied. Homoscedasticity, the fifth assumption is violated; thus, we report the robust standard errors instead, to treat this situation.

Several studies discuss that projects with videos tend to be more successful than those without videos. Our results imply the same effect. On the one hand, this could be a factor of the preparedness of a project as a whole. Not only having a video but also having a thoroughly and interestingly written description, the inclusion of visually pleasing images or other factors related to this would be expected to have a positive effect on the success of a campaign. We will discuss more the difference between causality and correlation in a section at the end of this chapter.

Hypotheses

The hypotheses presented in Chapter 4 can be tested using the model introduced in Section 6.1 and interpreted through the APE presented in Table 6.3.

Hypothesis 1

Increasing the amount of money asked for the project has a negative impact on the success of project funding.

By log-transforming the variable *goal_amount* we successfully reduce the variability of our data. The interpretation will be different from the one of level-level regression, which is relevant for our other variables. With other variables held constant, the effect of increasing the goal amount by 1% would be expected to lead to a decrease in the success of 0.073%. However, it is more conceivable to increase the goal amount by at least 10%. With that sort of increase, we would expect the chances of the project being successfully funded go down by 0.73%. Even though the variable is statistically significant

Table 6.3: Average Partial Effects

	<i>funded</i>	
	Logit	APE
has_video	0.464*** (0.143)	0.091*** (0.030)
short_camp	0.325** (0.156)	0.064** (0.031)
log(goal_amount)	-0.357*** (0.067)	-0.0733*** (0.000)
rewards	0.068*** (0.009)	0.013*** (0.002)
large_city	0.271** (0.110)	0.054** (0.022)
technology	0.170 (0.240)	0.033 (0.047)
arts	-0.003 (0.140)	-0.001 (0.028)
education	0.378** (0.161)	0.075** (0.032)
entertainment	0.303* (0.158)	0.060* (0.031)
community	0.211 (0.240)	0.042 (0.047)
miscellaneous	0.277 (0.206)	0.055 (0.041)
updates	0.289*** (0.023)	0.057*** (0.004)
qa	-0.057* (0.031)	-0.011* (0.006)
word_count	-0.001*** (0.0002)	-0.0002*** (0.0000)
Constant	2.322*** (0.764)	2.322*** (0.764)
Observations	1,871	1,871
Pseudo R ²	0.171	
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

at even 1% significance level, the magnitude of the yielded APE is negligible. Therefore, Hypothesis 1 is not supported.

Hypothesis 2

The number of rewards has a positive impact on the success of the project.

Even though the effect of adding different rewards to the project is minuscule, it is statistically significant at 1% significance level, and thus the hypothesis can be supported.

Hypothesis 3

Updating the project page has a positive impact on the success of project funding.

For Hypothesis 3, we confirmed the results of several other studies analysing the effect of frequent updates. In our analysis, we were unable to study updating with respect to time, but in its general form, this hypothesis can also be supported. The expected effect is that with an additional update, the probability of a project being successful increases by 5.7%, *ceteris paribus*.

Hypothesis 4

Answering questions relevant to the project has a positive impact on the success of the project.

In this case, the resulting average partial effect yielded a different sign than was predicted, although the variable is significant at a 10% significance level. Consequently, this hypothesis cannot be supported. The opposite sign of this effect could be explained. Possibly, when there is a need for substantial additional explanation, it could signal that the backers do not receive enough information from the project description or a video and thus, have to ask many questions in the questions and answers section.

Hypothesis 5

The length of a project description has a positive impact on the success of the project.

For Hypothesis 5, we expected a positive effect of the variable *word_count* on the probability of a project being successfully funded. The variable is statistically significant. Yet, we cannot conclude any significant effect as the yielded APE is minuscule.

Hypothesis 6

Having a project listed in the category technology has a positive impact on successful project funding.

Hypothesis 6 cannot be supported due to the insignificance of the variable at any significance level.

Hypothesis 7

The presence of a video has a positive effect on successful project funding.

As was discussed earlier in this section, the presence of a video in a project campaign is statistically significant and has the same sign as was expected. A project with a video increases the chance of the project being successfully funded by 9.1%. This result confirms the outcomes of other studies discussed in Chapter 3.

Hypothesis 8

The location of a founder within a large city has a positive impact on the success of the project.

Supported by the works of Agrawal et al. (2015; 2011) we can say that the effect of being located in places with higher population densities, such as the three largest cities in the Czech Republic, is positive at a 5% significance level. Thus, Hypothesis 8 can be supported.

Hypothesis 9

Shorter campaign durations have a positive impact on successful project funding.

By setting the period at 30 days, which is a duration offered by the platform HitHit, we increase our chances of the project being successfully funded by 6.4%. Hypothesis 9 can be supported at a 5% significance level.

Overall, most of the hypotheses tested are claimed to be statistically significant. Hypothesis 4 was not supported due to the unexpected sign of the effect, and Hypotheses 1, 5, and 6 were not supported due to both statistical and practical insignificance.

6.2 Regression results - the OLS model

For the OLS regression, a slightly different dependent variable was used. The variable *success* selected for this analysis measures the percentage of a project being funded. Even though we do not consider the point of reaching 100% as being successful, we suppose that the higher the percentage, the better. This holds even for projects which had already reached 100%. The regression of only the successful projects was brought out. However, the results did not yield any conclusive answers. The table summarizing the results of this regression can be found in Appendix B with reported robust errors. However, analysing only successful projects would mean analysing the effect on the excess success of a project. In our analysis, we do not analyse only factors that could increase the collected amount of money above the specified goal amount. With that being said, we proceeded to analyse both successful and unsuccessful projects.

It was necessary to satisfy or control for the needed assumptions in order to propose a model. The theoretical background of the methodology in Chapter 5 is applied.

We propose four different models and aim to find the best fitting one. In the case of OLS regression, one can use the R^2 value as a goodness-of-fit measure. First, we create a model with all of the dependent variables in their linear form. After that, when we test for the joint significance of variable *word_count* and its quadratic form, we include the squared variable into the model as well.

Second, to satisfy the no multicollinearity assumption, we run a Variance Inflation Factor (VIF) test. The indices of the test measure how much the variance of the estimated regression coefficient is increased due to collinearity (James et al. 2013). Unsurprisingly, we find a strong relationship between the variable *word_count* and its quadratic form. To treat for that situation, we use centering of the variable. Thus, for the models with both the linear and quadratic form of the variable, we apply their centered versions; *word_count_cnt* and *word_count_cntsq*. The results of the VIF test can be found in Appendix A.

The next issue that needed treatment was the homoscedasticity assumption. We run the Breusch-Pagan test, which was introduced in Chapter 5 for all of the four models. The test yields results with a p-value lower than 0.05 for all of the models; therefore, we are forced to apply the robust standard errors for valid inference. The Table 6.4 presents an overview of the results with reported robust standard errors.

Table 6.4: OLS regression with robust errors

	<i>success</i>			
	(1)	(2)	(3)	(4)
has_video	0.120*** (0.034)	0.121*** (0.034)	0.093*** (0.032)	0.093*** (0.033)
short_camp	0.146*** (0.048)	0.144*** (0.048)	0.119*** (0.042)	0.118*** (0.042)
log(goal_amount)	-0.083*** (0.017)	-0.082*** (0.017)	-0.156*** (0.020)	-0.154*** (0.020)
rewards	0.015*** (0.002)	0.015*** (0.002)	0.005*** (0.002)	0.005*** (0.002)
large_city	0.040 (0.027)	0.040 (0.027)	0.055** (0.023)	0.055** (0.023)
technology	0.049 (0.078)	0.048 (0.078)	0.034 (0.072)	0.034 (0.072)
arts	0.029 (0.036)	0.033 (0.035)	0.048 (0.031)	0.051* (0.031)
education	0.073* (0.042)	0.074* (0.042)	0.053 (0.037)	0.054 (0.037)
entertainment	0.017 (0.043)	0.015 (0.043)	0.026 (0.037)	0.025 (0.037)
community	-0.013 (0.055)	-0.011 (0.055)	-0.023 (0.050)	-0.023 (0.050)
miscellaneous	0.032 (0.051)	0.036 (0.051)	0.016 (0.047)	0.019 (0.048)
updates	0.045*** (0.004)	0.046*** (0.004)	0.033*** (0.005)	0.034*** (0.005)
qa	0.013 (0.009)	0.013 (0.009)	-0.007 (0.007)	-0.007 (0.007)
word_count	-0.0002*** (0.0001)	-	-0.0001*** (0.0001)	
word_count_cnt	-	-0.00004*** (0.00001)	-	-0.00003*** (0.00001)
word_count_cntsq		0.000** (0.000)	-	0.000* (0.000)
backers	-	-	0.002*** (0.0005)	0.002*** (0.0005)
Constant	0.633*** (0.089)	0.497*** (0.089)	0.963*** (0.097)	0.871*** (0.100)
Observations	1,871	1,871	1,871	1,871
R ²	0.164	0.166	0.358	0.359
Adjusted R ²	0.158	0.159	0.353	0.354
F Statistic	26.063*** (14; 1856)	24.607*** (15; 1855)	69.031*** (15; 1855)	64.938*** (16; 1854)

Note:

*p<0.1; **p<0.05; ***p<0.01

In the results of the regression, we add two models with the variable *backers*. We do not choose either of these two models for our final analysis, as there is a high chance of the independent variable and this dependent variable being defined by one another. This could also be the reason behind the inflated value of both R^2 and adjusted R^2 in comparison to the models without the variable *backers*.

Due to the problems arising with the inclusion of the *backers* variable and the joint significance of the variable *word_count* and its quadratic form, we focus on the model (2), given by the following Equation 6.2.

$$\begin{aligned}
 success = & \beta_0 + \beta_1 has_video + \beta_2 short_camp \\
 & + \beta_3 \log(goal_amount) + \beta_4 rewards + \beta_5 large_city \\
 & + \beta_6 technology + \beta_7 arts + \beta_8 education \\
 & + \beta_9 entertainment + \beta_{10} community + \beta_{11} miscellaneous \\
 & + \beta_{12} updates + \beta_{13} qa + \beta_{14} word_count_cnt \\
 & + \beta_{15} word_count_cntsq
 \end{aligned} \tag{6.2}$$

Additionally, another regression with the dependent variable *average_contribution* was applied and its results can be found in Appendix B. The model was abandoned after a subsequent consideration of the relevance of the dependent variable for the purpose of this thesis.

6.2.1 Interpretation of results

The results of the OLS regression of Model (2) are summarized in the second column of Table 6.4 on page 39. Considering the suggested hypotheses we do learn that most of the variables of our interest are statistically significant. We will exclude those that are statistically insignificant from any further commentaries. That includes variables *large_city*, *technology*, and *qa*. Therefore, the hypotheses concerning these variables cannot be supported.

From now on, we will focus on statistically significant variables. We notice that the only significant category dummy variable is *education*. Although it is significant at only a 10% significance level, we can claim that projects falling into the educational category are expected to raise 7.4% more of the goal amount than the other ones. People's engagement in educational activities could explain this. However, this should be a question about the motivation

of the backers rather than an influential factor from the creator's side. Since crowdfunding is a P2P tool, before we make any conclusions, we should be able to consider both sides of the 'contract.'

As was the case of the logit model, we see that the effect of increasing the goal amount by a relevant percentage, e.g., 10%, does not bring any practically significant values. In our case, increasing the goal amount by 10%, *ceteris paribus*, we would expect the collected money percentage to decrease by only 0.8%. All in all, we do not consider this to be a strong result. The same reasoning lies behind the rejection of the hypothesis concerning the description length. While both the variables, the linear and quadratic form, are individually and jointly statistically significant, the effect their respective estimates represent is negligible.

For the rest of the dependent variables, we do conclude both statistical and practical significance. All of these effects have the expected direction, and they do support the findings of studies applied to global datasets.

6.3 Causality vs Correlation

Before concluding our findings, we need to consider the difference between causality and correlation. Correlation tells us how strongly the pair of variables, dependent and independent, are linearly related and change together. It does not tell us how and why behind the relationship but only suggests that the relationship exists. While usually there are some clear examples, in the case of our analysis, detecting the difference between correlation and causality may be tricky.

The results of our analysis should be considered correlational, and we are left to speculate about the underlying mechanisms. In other words, the significance of the coefficients does not necessarily imply a causal effect on the outcome of the funding, for both the logit and OLS regression. The interpretation of causal effects is difficult, as some of the independent variables may proxy for factors we are unable to observe such as the throughout preparedness of a project, the creator's abilities and skills, or some unobserved activities of the creator. The outcomes of our research that are statistically significant may present opportunities for future research examining the exact associations between the characteristics of the project and higher funding success.

Chapter 7

Conclusion

The purpose of this thesis was to analyse the factors determining the success of a project funding of reward-based crowdfunding in the Czech Republic. In order to do that, several regressions were run to find the relationships between the characteristics of a project and its success, or even its success rate. This thesis was motivated by other studies that focused on crowdfunding in a global sense Agrawal et al. (2014); Mollick (2014); Belleflamme et al. (2010); Chan et al. (2018) analysing datasets from worldwide platforms such as Kickstarter.

First, a logit regression was used to determine the most influential factors and compare them with the results from those already mentioned studies from which we adopted most of our hypotheses. The average partial effects (APE) served to interpret the estimates of the regression. We confirmed that the presence of a video is connected to the success of a project. In particular, projects with videos are 9% more likely to obtain sufficient funding. Additionally, the results yielded a positive effect of frequent updating of a project, resulting in more than 5% higher probability of a project being successfully funded. Moreover, a similar positive effect holds for projects where the creator is based in one of the three largest cities in the Czech Republic.

The second applied regression was the OLS. In this case, we were studying the factors affecting the magnitude of the percentage of the money collected of the goal amount. Most of the results are similar to the ones of logistic regression. However, we may notice the insignificance of some variables which yielded a significant result in the logit model estimation. Since the OLS regression is not as popular as the logistic regression in studies concerning crowdfunding, we assume the results from the logistic regression to be more relevant. A factor that yielded similar and statistically significant estimates in both the regres-

sions is the length of a campaign. We confirm the hypothesis that projects with a shorter duration (in our case 30 days) are likely to be more successful than those with a longer duration.

Other factors with a positive effect on the crowdfunding success would also be the number of rewards and listing the project in the category *education*.

Some of the tested hypotheses could not be supported either due to practical or statistical insignificance. The most surprising was probably the almost null effect of the text length of a project description. Even though it was statistically significant, it did not yield any conclusive answers. In future research, the text of a project description could also be looked at qualitatively. While we looked only at the metrics of readability by counting the words, another view could be the count of syllables, sentences, and other measurable features of a text that give a more detailed insight.

Another practically insignificant result was the effect of the goal amount size. We supposed the effect to be strongly negative, while it yielded results that by increasing the goal amount by 10%, one would expect the probability of the project being funded to decrease by less than 1%. Possibly, the Czech project creators do not ask for unreasonably high amounts of money as much as creators in other countries.

For future research, I would also suggest enriching the dataset of a time factor. What I mean is tracking the activity of a project creator throughout time, studying the frequency of updates, questions, ideally even Facebook or other social network activity. A dataset with the creators' characteristics would also help to get more insight into what lies behind the success of individual projects.

Bibliography

- Agrawal, A., Catalini, C. and Goldfarb, A. (2011), 'The Geography of Crowdfunding.', *NBER Working Papers* .
- Agrawal, A., Catalini, C. and Goldfarb, A. (2014), 'Some Simple Economics of Crowdfunding.', *Innovation Policy and the Economy* p. 63.
- Agrawal, A., Catalini, C. and Goldfarb, A. (2015), 'Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions', *Journal of Economics & Management Strategy* **24**(2), 253–274.
- Belleflamme, P., Lambert, T. and Schwienbacher, A. (2010), Crowdfunding: An Industrial Organization Perspective, *in* 'Prepared for the workshop Digital Business Models: Understanding Strategies', Citeseer, pp. 25–26.
- Belleflamme, P., Lambert, T. and Schwienbacher, A. (2014), 'Crowdfunding: Tapping the Right Crowd.', *Journal of Business Venturing* **29**(5), 585–609.
- Bi, S., Liu, Z. and Usman, K. (2017), 'The Influence of Online Information on Investing Decisions of Reward-based Crowdfunding', *Journal of Business Research* **71**, 10–18.
- Brabham, D. C. (2013), *Crowdsourcing.*, The MIT Press Essential Knowledge Series, The MIT Press.
- Burtch, G., Ghose, A. and Wattal, S. (2013), 'An Empirical Examination of the Antecedents and Consequences of Contribution Patterns in Crowd-funded Markets', *Information Systems Research* **24**(3), 499–519.
- Chan, C. S., Park, H., Patel, P. and Gomulya, D. (2018), 'Reward-based Crowdfunding Success: Decomposition of the Project, Product Category, Entrepreneur, and Location Effects', *Venture Capital* .

Crowdfunding by Country (2017).

URL: <http://eurocrowd.org/crowdfunding-by-country/>

De Buysere, K., Gajda, O., Kleverlaan, R., Marom, D. and Klaes, M. (2012), ‘A Framework for European Crowdfunding’, *A framework for European crowdfunding* .

Dingman, S. (2013), ‘Canadian’s Smartwatch Startup Matches Record 15 Million Dollars in VC Funding’.

URL: <https://www.theglobeandmail.com/technology/business-technology/canadians-smartwatch-startup-matches-record-15-million-in-vc-funding/article11965214/>

Ellenoff, D. S. (2017), Review of Crowdfunding Regulation 2017. Interpretations of Existing Regulation Concerning Crowdfunding in Europe, North America and Israel, Technical report, The European Crowdfunding Network AISBL.

Ennico, C. (2016), *The Crowdfunding Handbook*, Grupo Nelson.

GoFundMe (2018), ‘Choosing a campaign type’, *GoFundMe - support* .

URL: <https://support.gofundme.com/hc/en-us/articles/203603984-Choosing-a-Campaign-Type>

Gorbatai, A. D. and Nelson, L. (2015), ‘Gender and the Language of Crowdfunding’, *Academy of Management Proceedings* **2015**(1), 15785.

Greenberg, M. D., Pardo, B., Hariharan, K. and Gerber, E. (2013), Crowdfunding Support Tools: Predicting Success & Failure, in ‘CHI ’13 Extended Abstracts on Human Factors in Computing Systems’, CHI EA ’13, ACM, New York, NY, USA, pp. 1815–1820.

HitHit.cz - FAQ (2019).

URL: <https://www.hithit.com/cs/article/faq>

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013), *An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics Book 103)*, Springer.

Kleemann, F., Voss, G. and Rieder, K. (2008), ‘Un(der)paid Innovators: The Commercial Utilization of Consumer Work Through Crowdsourcing’, *Science, Technology & Innovation Studies (STI-Studies)* **4**.

- Koning, R. and Model, J. (2013), 'When Nothing Is Better Than Something: Cognitive Status and the Hazard of Small Early Investments', *SSRN Electronic Journal*.
- Massolution (2015), *2015CF Crowdfunding Industry Report*, Massolution.com.
- Mitra, T. and Gilbert, E. (2014), The Language that Gets People to Give: Phrases that Predict Success on Kickstarter, in 'Phrases that predict success on Kickstarter', pp. 49–61.
- Mollick, E. (2014), 'The Dynamics of Crowdfunding: An Exploratory Study?', *Journal of Business Venturing* **29**(1), 1 – 16.
- Morduch, J. (1999), 'The Microfinance Promise', *Journal of Economic Literature* **37**(4), 1569–1614.
- Sauermann, H., Franzoni, C. and Shafi, K. (2019), 'Crowdfunding Scientific Research: Descriptive Insights and Correlates of Funding Success', *PLOS ONE* **14**(1), e0208384.
- Stegmaier, J. (2015), *A Crowdfunder's Strategy Guide: Build a Better Business by Building Community*, Berrett-Koehler Publishers.
- Steinberg, S. (2012), *The Crowdfunding Bible: How to Raise Money for Any Startup, Video Game or Projects*, Portfolio.com.
- Wooldridge, J. M. (2015), *Introductory econometrics: A modern approach*, Nelson Education.
- Young, T. E. (2012), *The Everything Guide to Crowdfunding*, Simon + Schuster Inc.
- Ziegler, T., Shneor, R., Wenzlaff, K., Odorovic, A., Johanson, D., Hao, R. and Ryll, L. (2019), *Shifting Paradigms: The 4th European Alternative Finance Benchmarking Report*, Cambridge Centre for Alternative Finance.

Appendix A

Tests and Assumptions

Table A.1: LPM model - robustness check for logit

	<i>Dependent variable:</i>
	funded
has_video	0.105*** (0.028)
short_camp	0.077** (0.031)
log(goal_amount)	-0.073*** (0.013)
rewards	0.013*** (0.002)
large_city	0.053** (0.022)
technology	0.042 (0.048)
arts	0.003 (0.028)
education	0.081** (0.032)
entertainment	0.064** (0.032)
community	0.041 (0.048)
miscellaneous	0.057 (0.042)
updates	0.041*** (0.003)
qa	-0.007 (0.006)
word_count	-0.0002*** (0.00005)
Constant	0.481*** (0.070)
Observations	1,871
R ²	0.190
Adjusted R ²	0.184
Residual Std. Error	0.452 (df = 1856)
F Statistic	31.094*** (df = 14; 1856)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table A.2: Variance Inflation Factors

	VIF1	VIF2	VIF3	VIF4
has_video	1.093	1.093	1.095	1.095
short_camp	1.011	1.011	1.012	1.012
log(goal_amount)	1.167	1.170	1.219	1.223
rewards	1.199	1.199	1.273	1.273
large_city	1.031	1.031	1.032	1.032
TECH	1.198	1.198	1.198	1.198
ARTS	1.374	1.379	1.375	1.380
EDUC	1.942	1.942	1.943	1.943
ENTR	2.306	2.307	2.306	2.307
COMM	1.332	1.332	1.332	1.332
MISC	1.325	1.327	1.325	1.328
updates	1.145	1.147	1.175	1.177
qa	1.107	1.107	1.127	1.127
word_count_cnt	-	1.491	-	1.497
word_count_cntsq	-	1.361	-	1.362
backers	-	-	1.095	1.270

The Gauss-Markov Assumptions

Under the assumptions MLR.1 through MLR.5, the $\hat{\beta}_j$ estimator is the best linear unbiased estimator (BLUE) (Wooldridge 2015).

MLR.1. Linear in Parameters The equation of the model is always linear in parameters.

MLR.2. Random Sampling We have a random sample of n observations following the population model from Assumption MLR.1.

MLR.3. No Perfect Collinearity None of the independent variables is constant and there are no exact linear relationships among them.

MLR.4. Zero Conditional Mean The error term u has an expected value of zero given any independent variables.

MLR.5. Homoskedasticity The error term u has the same variance given any values of the explanatory variables.

Appendix B

Additional Regressions

Table B.1: OLS regression - only successful projects

	<i>Dependent variable:</i>	
	success	
	(1)	(2)
has_video	-0.008 (0.033)	-0.0004 (0.031)
short_camp	0.059 (0.056)	0.059 (0.052)
goal_amount	0.008 (0.018)	-0.089** (0.040)
rewards	0.001 (0.001)	-0.002* (0.001)
large_city	-0.024 (0.026)	-0.010 (0.023)
technology	0.044 (0.118)	0.046 (0.116)
arts	0.058 (0.038)	0.068* (0.036)
education	-0.043 (0.052)	-0.029 (0.045)
entertainment	-0.110** (0.052)	-0.087* (0.047)
community	-0.113** (0.051)	-0.112** (0.047)
miscellaneous	-0.072 (0.055)	-0.067 (0.053)
updates	-0.002 (0.004)	-0.001 (0.003)
qa	0.037*** (0.011)	0.020*** (0.008)
word_count_cnt	0.00000 (0.00001)	0.00000 (0.00001)
word_count_entsq	0.000 (0.000)	0.000 (0.000)
backers		0.001** (0.0003)
Constant	1.188*** (0.091)	1.515*** (0.154)
Observations	954	954
R ²	0.062	0.179
Adjusted R ²	0.047	0.165
Residual Std. Error	0.418 (df = 938)	0.391 (df = 937)
F Statistic	4.123*** (df = 15; 938)	12.795*** (df = 16; 937)

Note:

*p<0.1; **p<0.05; ***p<0.01

Table B.2: OLS regression - *average_contribution*

	<i>Dependent variable:</i>
	average_contribution
has_video	81.356 (52.085)
short_camp	55.617 (57.089)
log(goal_amount)	241.899*** (24.112)
rewards	-3.196 (2.753)
large_city	79.619** (40.235)
technology	168.269* (87.845)
arts	52.196 (51.408)
education	-64.436 (59.101)
entertainment	27.812 (58.246)
community	-71.875 (87.405)
miscellaneous	-163.869** (76.446)
updates	8.168 (5.213)
qa	-16.062 (10.726)
word_count_cnt	-0.007 (0.014)
word_count_entsq	0.00000 (0.00000)
Constant	-293.772** (127.645)
Observations	1,871
R ²	0.075
Adjusted R ²	0.067
Residual Std. Error	827.075 (df = 1855)
F Statistic	9.956*** (df = 15; 1855)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01