Signaling to Overcome Informational Asymmetry: An Analysis of the Determinants of Crowdfunding Success of Technology Start-Ups on Kickstarter

STA302: Methods in Data Analysis I

Fall 2024

Note: This paper is adapted from an analysis of the determinants of crowdfunding success of start-ups on Kickstarter I did with Ryan Junejo for a statistics course. I rewrote the introduction and conclusion to provide more context for the project and I condensed the methods and results sections for brevity.

Introduction

The rise of the internet has enabled new forms of capital fundraising where founders collect small sums of money from their social networks and small investors through crowdfunding. Many successful projects, especially in the technology sector, have arisen from crowdfunding, notably: three-dimensional printers, virtual reality glasses, and smart watches. The largest crowdfunding platform, Kickstarter, was launched in 2009 and has received over $7 billion dollars in pledges from over twenty million backers as of 2023.

The rise of crowdfunding has been excitedly anticipated by many in the business community specifically for its potential to offer a source of seed funding for non-North American start-ups where a lack of angel investors has often been a challenge.

Crowdfunding fills a gap in the financing system by providing an opportunity for more efficient capital allocation for some types of start-ups. Grüner and Siemrooth (2019) showed using a Bayesian investment game that the most efficient capital allocation occurs when there is minimal wealth inequality between investors (otherwise investments reflect preferences of wealthy investors). Thus, traditional capital investment reflects only the preferences of those with sufficient wealth to invest; whereas crowdfunding is able to provide extra information about the market for products. .

In this vein, crowdfunding can be viewed as a sort of "prepurchasing" providing valuable information about the demand for the product being offered. Unlike market analysis where individuals claim to be willing to purchase a product, consumers are actually willing to provide capital to prepurchase it.

Furthermore, crowdfunding can reduce the risk to the entrepreneur in an all-or-nothing model because if demand is insufficient such that they fail to cover costs, then the project will not proceed (Cumming et al., 2016). This reduces the risk to the entrepreneur because no financial resources have yet been used.

That said, crowdfunding markets come with several downsides to traditional business financing avenues. During traditional investing, if an entrepreneur fails to receive funding during their first attempt they can change their business proposal and approach another backer. Crowdfunding is facilitated through the Internet where information about a failed campaign remains. In practice, this means that entrepreneurs almost never get a second chance to launch a successful campaign. This precludes the possibilities for entrepreneurs to revise and improve their business plan after an initial rejection.

Crowdfunding also faces unique challenges of ensuring the "quality". Crowdfunding platforms like Kickstarter have an incentive to minimize such transactions to prevent a collapse of the market.

Informational Asymmetries in Crowdfunding Markets

Problems with informational asymmetries are more heightened in crowdfunding than other entrepreneurial finance markets. Unlike larger investors, crowdfunders have neither the ability nor the incentive to devote substantial amounts of energy to due diligence. This is due both to an informational asymmetry and the fact that the small amounts invested removes the incentive for any individual crowdfunder to devote substantial effort to due diligence.

This creates a collective action problem where it would be most efficient for due diligence to be done on investments, however no individual has the incentive to perform proper due diligence, and there is no mechanism to coordinate who pays for due diligence.

The lack of due diligence in crowdfunding platforms may lead to a hesitancy of buyers to use it and ultimately a devaluing of investment that could cause an Akerlof-like (1978) collapse of the crowdfunding market.

This collapse has not occurred. In fact, Iyer et al. (2015) show that the crowd seems able to assess the risk of a project and predict failure at least as well as traditional bankers.

Signals play a crucial role. An effective signal is one in which it is more costly for the low-value start-ups to signal than high value start-ups, creating a separating equilibrium. Stiglitz (2000) suggests two main types of informational asymmetry of particular importance: information about quality and about intent. Signaling about intent is only useful when there are mechanisms to penalize entrepreneurs who do not follow through on their promises.

Reward-based projects and donation-based projects require different signals. Many users engage to support a project they like rather than as an investment. The goal of this paper is to not to examine charitable projects, but rather to characterize the determinants of crowdfunding success for reward-based technology start-ups on Kickstarter.

Methods & Results

Exploratory Data Analysis

The durations were calculated as the difference between the end date and launch date for each Kickstarter campaign and were grouped into three categories: *short* (1–15 days), *medium* (16–30 days), and *long* (31–60 days). The category predictor was regrouped into three groups: *Hardware, Software & Web*, and *Other Technology*, to focus on broader industry trends.

Several predictors, specifically **usd_goal_real** and **usd_pledged_real**, were right-skewed. Box-Cox transformations were applied to correct these distributions as seen in *Figure 1*. Histograms confirmed that these transformations approximated normality. The residual plots indicated no violations for categorical predictors (**category_group, is_North_America**, and **duration_cat**) so they were left untransformed post-grouping.

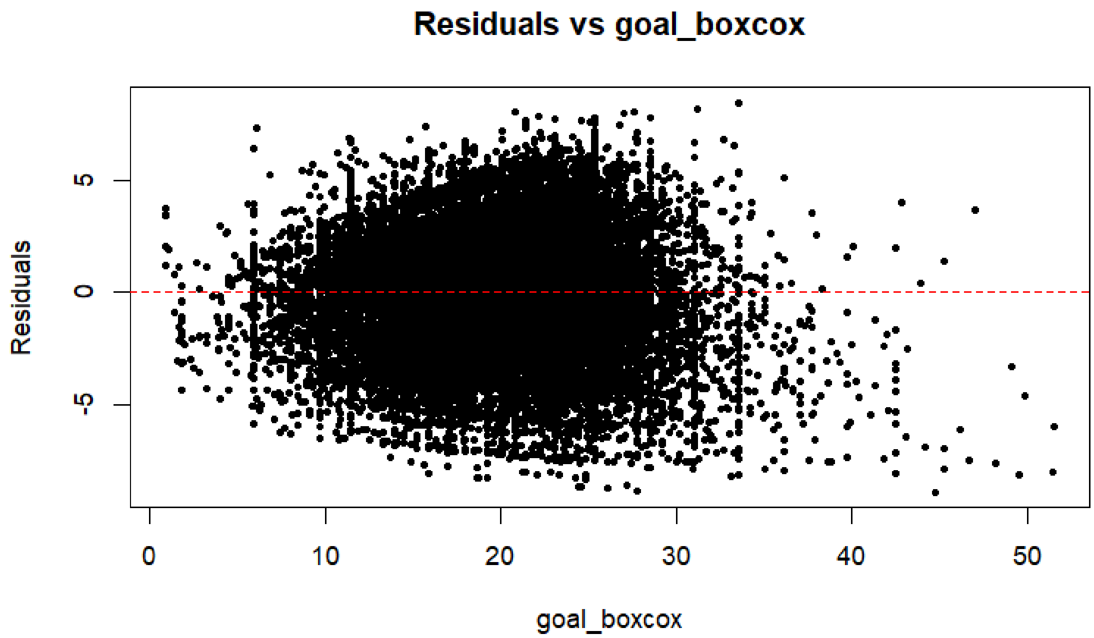

Figure 1. Residuals plotted against the goal predictor after a Box Cox transformation, with a horizontal reference line at zero to assess homoscedasticity.

Initially, the potential predictor backers, however no stable transformation could address its significant skewness and deviations from normality, so it was excluded from the final model.

Conditional mean assumptions were checked in Figure 2 using scatterplots of the response variable versus fitted values, confirming a mostly random scatter with no identifiable nonlinear trends. Pairwise scatterplots of predictors in *Figure 3* revealed no obvious curves or other violations of the conditional mean assumptions for linear regression.

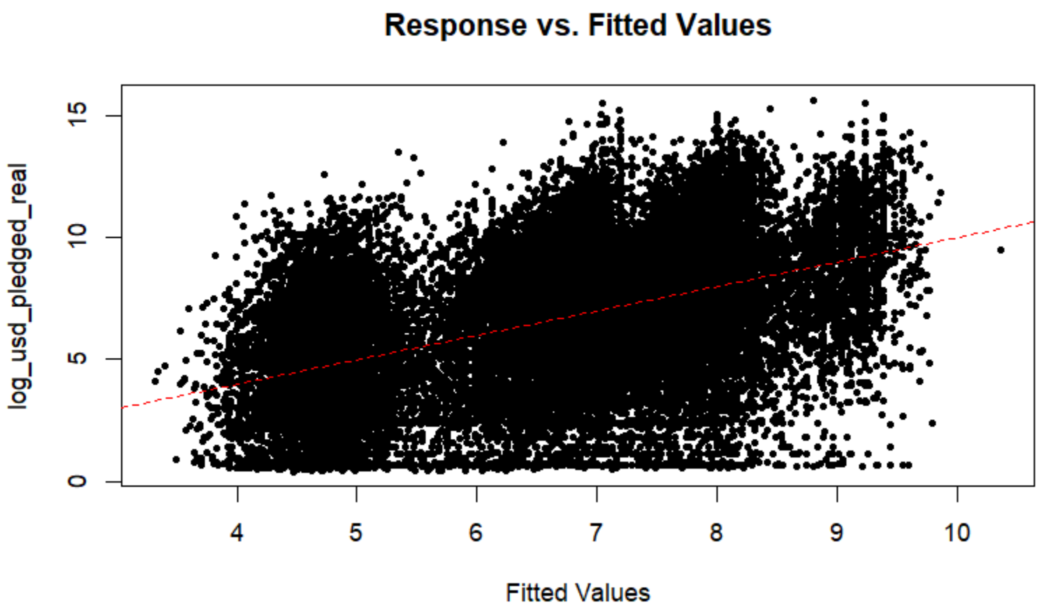

Figure 2. Scatterplot of Response versus Fitted values

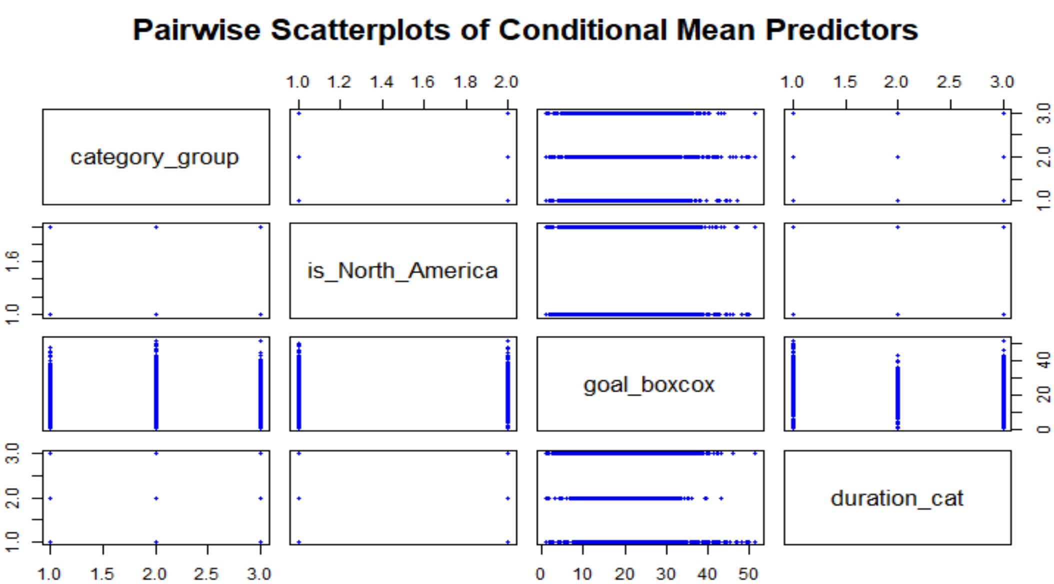

Figure 3. All Pairwise Scatterplots of Predictors

Model Diagnostics

Residual Diagnostics

Residual diagnostics were conducted to evaluate the model's accuracy and assumptions. A residual vs. fitted plot indicated a constant variance violation. However, since we had already selected the best transformation available, we could not further address this issue without compromising other aspects of the model. Despite this limitation, the plot generally supported the assumption of homoscedasticity. A Q-Q plot) revealed slight deviations at the right tail but showed no major violations of the normality assumption as seen in (*Figure 5,* Appendix).

Multicollinearity

Variance Inflation Factor (VIF) values for all predictors were below 2, confirming no significant multicollinearity:

Table 1: Variance Inflation Factor (VIF) for Model Predictors

Predictor	**VIF**
category_group	1.02
is_North_America	1.01
goal_boxcox	1.03
duration_cat	1.04

Problematic Points

Problematic points were inspected using the hat matrix, Cook's distance (cutoff: 0.9180), DFFITS (cutoff: 0.0344), and DFBETAS (cutoff:0.1217) Among these diagnostics, 3% of observations were flagged as influential based on DFFITS. Using the Cook's cutoff distance, it was observed that their retention did not significantly affect model fit or coefficients, and no points were removed.

Model Performance

An ANOVA was conducted to identify the existence of a linear relationship. We thus concluded a statistically significant linear relationship exists with at least one predictor because the F-statistic of 1,255 confirmed the model's overall significance. with p-value ($p < 2.2 \times 10$^−16). We then performed t-tests to check the significance ($p < 0.05$) of each predictor in the model and found all predictors to have statistically significant contributions to the model and their effects aligned with expectations from the literature.

Table 2: T-Test Results for Predictor Significance in Kickstarter Campaign Model

Predictor	Sum of Squares	Mean Square	F-Value	p-Value
category_group	46,881	23,440.7	3108.97	$< 2 \times 10^{-16}$
is_North_America	463	462.9	61.39	4.847×10^{-15}
goal_boxcox	3,894	3,893.6	516.41	$< 2 \times 10^{-16}$
duration_cat	5,551	2,775.5	368.12	$< 2 \times 10^{-16}$

The final model explained 21.81% of the variance in the log-transformed pledged amounts, with the $R^2 = 0.2181$ and $adjusted\ R^2 = 0.2179$, as seen in Table 3. The low R^2 and R^2_{adj} values reflect the inherent variability of crowdfunding outcomes, yet the predictors provide meaningful insights into factors influencing pledged amounts. Using stepwise selection via StepAIC, the analysis confirmed that no model with a subset of the predictors attained a lower AIC, and thus no predictors were removed.

Model Validation

The final model was validated using a 70-30 train-test split, and performance metrics were assessed to confirm generalizability. The minimal differences between the training and testing metrics (a difference of 0.222 for MSE and -0.0066 for R^2) indicate that the model does not overfit and performs satisfactorily on unseen data.

Conclusion

Informational asymmetry or the problem of "lemons" presents a crucial challenge to the crowdfunding market. Without effective signals, the market risks facing a market collapse. My model of crowdfunding success gives a sense of several signals that exist for technology start-ups on the platform Kickstarter in determining the amount pledged.

I find that about 22% of variation in amount pledged to technology start-ups on Kickstarter can be explained through the category (Hardware, Software & Web, or Other); and if it is North American, the goal and the duration of the campaign.

In countries where entrepreneurs lack sufficient seed capital, crowdfunding has been celebrated for its potential to offer a pathway to raise funding for those in non-North American contexts where there lacks angel investors for initial seed funding (Hornuf and Schwienbacher 2017). My findings suggest such optimism should be tempered: significant discrepancies in funding success between North American and non-North American crowdfunding start-ups persist.

My model confirms the finding in Cumming et al. (2016) that the crowd responds to the funding goal. It also shows that hardware products tend to be more successful than software and other technology products, with software and web being the least successful. This likely reflects the fact that hardware projects offer a tangible product.

Likewise, the length of a campaign appears to be important to the success of a crowdfunding campaign. Vismara (2017) shows that early contributions are fundamental to the success of a crowdfunding campaign and offer a signal to overcome informational asymmetry in crowdfunding. The positive relationship between medium campaigns over long campaigns likely reflects the success of start-ups that do well in early fundraising. A long duration campaign is also perhaps a signal suggesting a lack of confidence.

This study looked at existing signals in the crowdfunding market. Not all signals are effective signals at differentiating low-quality start-ups from high quality start-ups. While this model looks at the determinants of a successful fundraising campaign, it does not examine the ultimate success of these ventures. More research is needed to examine determinants of successful technology ventures on Kickstarter.

References

Akerlof, G. A. (1978). The market for “lemons”: Quality uncertainty and the market mechanism. In *Uncertainty in economics* (pp. 235-251). Academic Press.

Grüner, H. P., & Siemroth, C. (2019). Crowdfunding, Efficiency, and Inequality. *Journal of the European Economic Association*, *17*(5), 1393–1427. https://doi.org/10.1093/jeea/jvy023

Hornuf, L., & Schwienbacher, A. (2017). Should securities regulation promote equity crowdfunding? Small Business Economics, 49(3), 579–593. https://doi.org/10.1007/s11187-017-9839-9

M. Mouillé (2019). Kickstarter Projects Dataset, Kaggle, [Online]. Available: https://www.kaggle.com/datasets/kemical/kickstarter-projects

Appendix

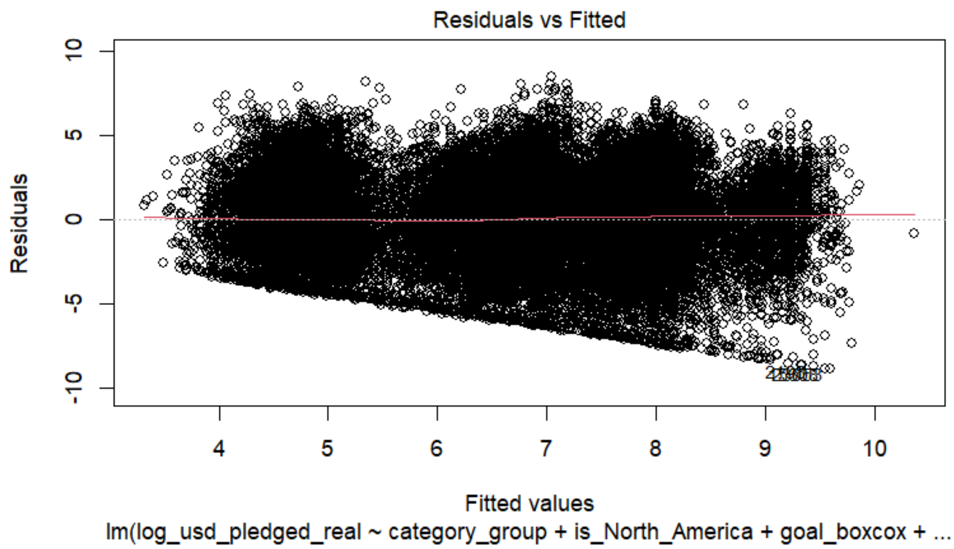

Figure 4. Residuals vs Fitted Values of the linear model.

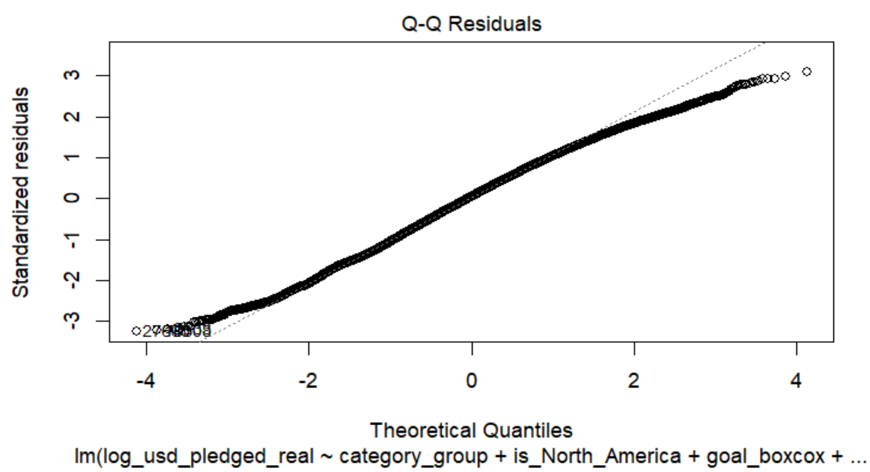

Figure 5: Normal Q-Q Plot: Evaluating the normality of residuals

Table 3: Linear Regression Model Results for Kickstarter Campaign Pledges

Predictor	**Estimate**	**Std. Error**	**t-Value**	**p-Value**	**Interpretation**
Intercept	6.557	0.082	79.79	< 2 x 10^-16	Baseline log-pledged amount for campaigns in North America.

Predictor	**Estimate**	**Std. Error**	**t-Value**	**p-Value**	**Interpretation**
category_group (Other Tech.)	-1.082	0.042	-25.85	< 2 x 10^-16	Pledged amounts are lower for "Other Tech." campaigns.
category_group (Software & Web)	-2.907	0.040	-72.67	< 2 x 10^-16	Campaigns in "Software & Web" yield the lowest pledged amounts.
is_North_America (Not NA)	-0.307	0.038	-8.08	6.79 x 10^-16	Non-North American campaigns receive lower pledged amounts.
goal_boxcox	0.0680	0.003	20.35	< 2 x 10^-16	Higher goals are positively associated with pledged amounts.
duration_cat (Medium)	1.112	0.053	21.12	< 2 x 10^-16	Medium-duration campaigns attract higher pledged amounts.
duration_cat (Short)	-0.121	0.041	-2.95	0.003	Short campaigns are slightly less successful in raising funds.