



4-2015

Crowdfunding Success: The Short Story - Analyzing the Mix of Crowdfunded Ventures

Brendon Lichtig
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/wharton_research_scholars



Part of the [Business Commons](#)

Lichtig, Brendon, "Crowdfunding Success: The Short Story - Analyzing the Mix of Crowdfunded Ventures" (2015). *Wharton Research Scholars*. 121.

http://repository.upenn.edu/wharton_research_scholars/121

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/wharton_research_scholars/121

For more information, please contact repository@pobox.upenn.edu.

Crowdfunding Success: The Short Story - Analyzing the Mix of Crowdfunded Ventures

Keywords

crowdfunding

Disciplines

Business

CROWDFUNDING SUCCESS: THE SHORT STORY

- Analyzing the Mix of Crowdfunded Ventures -

Brendon Lichtig

Contents

Executive Summary.....	3
Motivation.....	4
Research Question	4
Background	4
Dataset.....	7
Overview.....	7
Representativeness	9
Methodology	10
Approach	10
Selected Model	14
Results	15
Conclusion.....	17

Executive Summary

In this paper, I create a definition of crowdfunding success that acknowledges project creators' variety of objectives and build a model to start understanding why some projects perform better than others. This is an extension of the research already conducted, which focuses on predicting funding completion.

Analysis is conducted on a sample of Kickstarter data from 2014 and relies on Empirical Bayes models, specifically the negative binomial distribution. The best model only uses three parameters and achieves a mean average percent error of 8%. The quality of the model, considering its simplicity and the fact that the data were not cleaned, is noteworthy.

My model tells a story of a heterogeneous mix of projects, whose success is affected by the number of projects the creator has previously launched. The importance of previously launched projects might reflect that creators improve with experience, or that prolific creators are able to build a following, or that only innately skilled people launch many projects. Additional data is required in order to build a case of causality around one of these forces.

Motivation

Research Question

Statistics-based crowdfunding research has thus far focused on predicting the funding success of projects. While the prediction models perform well, substantially all are designed to be black boxes. For example, Ph.D. candidate Vincent Etter et al. says “For now, our predictors only output a probability of success, but act as a black box: no reason for the probable success / failure is given” in “Launch Hard or Go Home!: Predicting the Success of Kickstarter Campaigns.” A logical next step is to better understand the driving factors, why some projects are more successful than others. This led me to my research question:

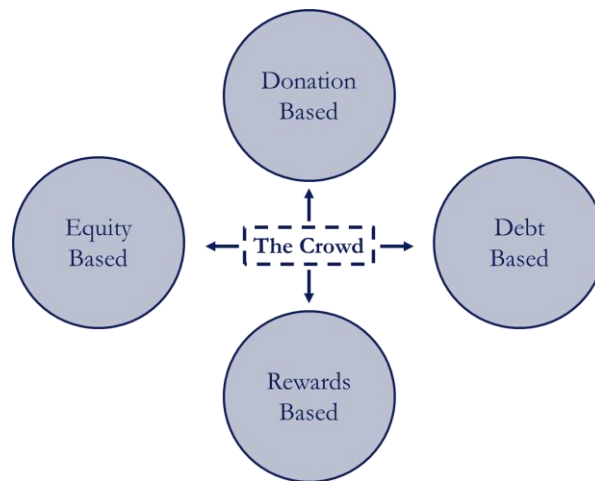
How can we empirically characterize the mix of crowdfunding projects’ success?

In this paper, I will lay the groundwork for a better understanding of what actually makes crowdfunding projects successful. Project creators would find this information particularly useful, as they contemplate using crowdfunding and how to go about launching.

Background

Starting at the ground level, crowdfunding is a special type of crowdsourcing, using (“sourcing”) a group (a “crowd”) of people to collectively achieve a task. In the case of crowdfunding, the task is to pool money (“funds”) for a given use, ranging from cancer research to getting a startup off the ground.

There are four main types of crowdfunding: donation-based, reward-based, equity-based, and debt crowdfunding. Equity-based is currently limited in the United States, as regulation prohibits many potential investors from participating. At present, it falls under Regulation D of the Securities Act of 1933, which limits equity crowdfunding to accredited investors, who need to have annual income greater than \$200,000 or over \$1MM in net worth (“Regulation Crowdfunding”).



The United States government recognizes crowdfunding is an increasingly popular channel for entrepreneurs and has thus mandated, via Title III of the JOBS Act, the creation of Regulation Crowdfunding. It has been designed to “provide a framework for the regulation of registered funding portals and brokers” (“Regulation Crowdfunding”) and make equity crowdfunding a viable option for small businesses. The regulation was presented as a proposed rule on October 23, 2013 and the comment period ended February 3, 2014. The rule is expected to be implemented in the latter part of 2015 or early in 2016.

Once the regulation is implemented, equity crowdfunding in the United States may develop into a more legitimate alternative to seed funding for entrepreneurs. If that happens, entrepreneurs will need to make difficult decisions regarding how to pursue funding. The race is against the clock to understand crowdfunding dynamics as best as possible before the regulation is rolled out.

Early research focused on survey and case-based explorations of crowdfunding dynamics. The subsequent wave involved researchers scraping data from dominant crowdfunding platforms and running a variety of prediction models, including logistic regressions, markov chains, and fixed effect models. Dr. Ethan Mollick’s “The Dynamics of Crowdfunding: An Exploratory Study” was one of the first of this kind in the United States with significant statistical power. Kickstarter is one of the most prominent donation-based platforms and has become the basis of Dr. Mollick’s empirical research, as well as that of many others in the research community.

I would like to highlight “Inferring the Impacts of Social Media on Crowdfunding” by Chun-Ta La, Sihong Xie, Xiangnan Kong, and Philip S. Yu because it currently sits on the frontier of predicting crowdfunding success. The team developed a model that is able to predict success with 75.7% accuracy using only information available after 5% of the project duration (“Interring the Impacts”). The model I develop will be of a different flavor, attempting to gain understanding rather than predictive power.

Dataset

Overview

I chose to focus on Kickstarter data, as it appears to be the most commonly used among researchers and the most accessible. The dataset was pre-scraped by the operators of thekickbackmachine.com, an online research tool, and contains information on 4,682 projects launched during April 2014. A total of 369 projects had 0 backers, which is unusual, so I excluded those from my analysis. While I could have made additional adjustments, I used the dirty data, to ensure the result would be as robust as possible.

Each project has 27 main variables available. I chose to limit the analysis to the five independent variables that I most strongly believed, ex ante, could influence success in a meaningful way. They are:

- Duration in days – how many days a project is live, open for contribution
- Creator, # projects created – the number of projects the creator has launched. This includes the current project, so the variable's minimum value is one
- Creator, # projects backed – the number of projects to which the creator has contributed
- Goal – the project's dollar value funding goal. On Kickstarter, if the project does not receive enough pledges by the end of the term, the creator gets \$0, so successful funding is all or nothing

In the analysis, I will use "Backers," the number of people who contributed to a given project, as the dependent variable. The reasoning behind this decision is discussed in the "Methodology" section.

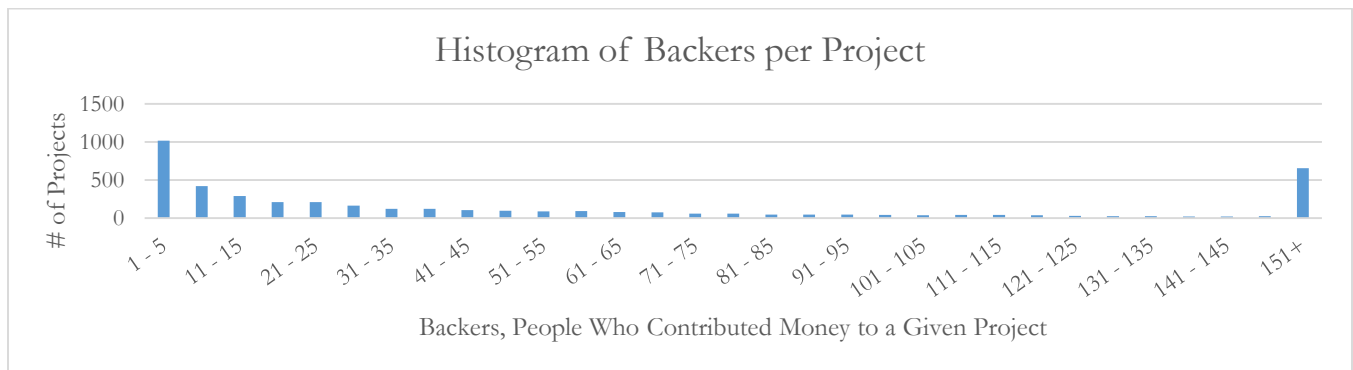
Here is a sample of the data:

Project #	Backers	Duration in Days	Creator, # Projects Crtd.	Creator, # Projects Bckd.	Goal
1	23	30	1	0	\$115,000
2	1195	30	3	2	\$29,995
3	3	31	1	13	\$5,000
4	4	45	1	2	\$15,000
5	4	31	1	0	\$5,500
6	206	30	3	42	\$20,000
7	5	60	1	0	\$20,000
8	1	40	1	2	\$20,000
9	25	35	1	0	\$50,000

Below are summary statistics of the 4,313 projects I ultimately used in modeling:

Projects:	
Count	4313
% of Total	92%
Backers:	
Min	1
Max	23,999
Mean	140
Median	26

Below is a visualization of the data:



Note the long tail and concentration of projects with relatively few backers. The tail extends to one project that garnered 23,999 backers and most of the action dies at >2,000, yet 80% of the projects accumulated less than 100 backers.

The histogram groups projects by number of backers in increments of five and a final bucket is formed by projects with 150+ backers. In choosing the bucket size, I tried to balance keeping the groups small enough to retain a meaningful interpretation (getting 21 backers is not substantially different than getting 25) and large enough to eliminate some of the noise that disrupts the monotonically decreasing pattern we would expect to see. In choosing the number of buckets, I wanted to ensure the final 150+ bucket was not excessively large but that the number of projects per bucket did not become too low.

Representativeness

I would like to apply the model's results to equity crowdfunding in the United States after Regulation Crowdfunding is implemented, however, the following are potential concerns with the representativeness of the data:

- Since Kickstarter success is all or nothing, it may have different dynamics from equity crowdfunding, depending on how the platform is structured
- The data is only from April, 2014. While I am not aware of any distorting influences, a sample of projects over a number of months would have been preferred
- While there is overlap, the mix of projects on Kickstarter versus equity crowdfunding platforms differs. For example, equity crowdfunding seems to have a higher concentration of technology ventures based on my observations
- Similarly, the mix of investors on Kickstarter may not be fully representative of equity crowdfunding investors, as the reward structures attract people with differing motives. That said, I would expect for there to be a good amount of overlap, constituted by professionals in the entrepreneurship / venture financing space

Methodology

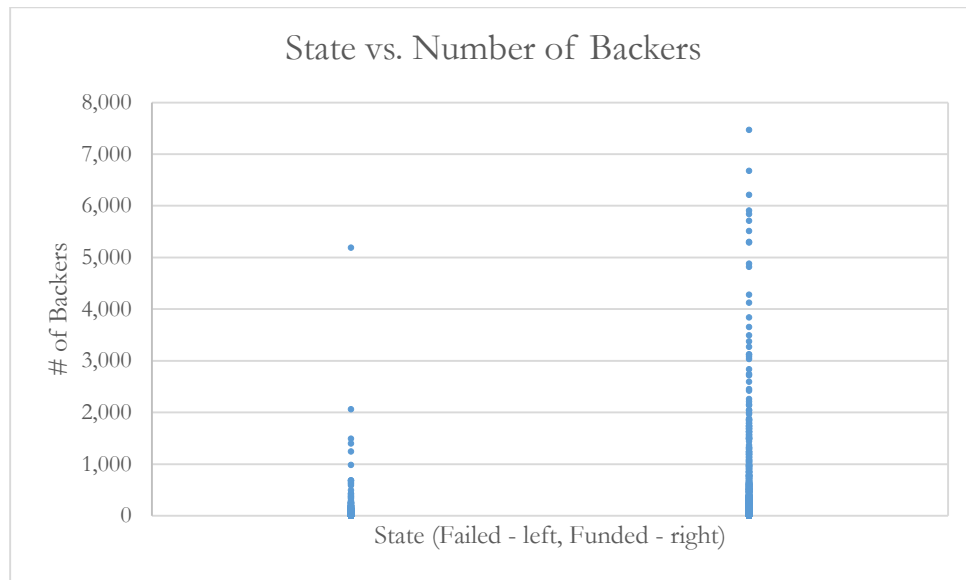
Approach

Definition of Success

Project creators can be drawn to Kickstarter for a variety of reasons, ranging from a pure desire for funding to capitalizing on the marketing opportunity. With multiple objectives in play, the definition of success varies from project to project. The easy metric to lean on is the project's ultimate state: successfully funded or failed. On Kickstarter, that is determined solely based on if the project met its funding goal.

While research thus far has focused on funding completion, I chose to focus on the number of backers. Unlike funding completion, the number of backers produces a gradient of success – the more backers, the more successful the project both in terms of funding and marketing. Below, a table and plot suggest that my success metric maps fairly well into that of previous researchers:

	% of Failed	% of Funded
Backers	Projects	Projects
>1000	0%	5%
>500	1%	10%
>200	2%	24%
>100	6%	42%
>50	13%	65%
>25	24%	83%
>10	43%	96%



Only 24% of projects that fail have more than 25 backers. Conversely, only 17% of successful projects had less than 25 backers. Thus successfully funded projects tend to have many backers and vice versa.

Average contribution size is an important consideration, and constitutes the missing half of this analysis. I unfortunately do not have access to adequately granular contribution data, so I have included that in the discussion of future analysis under "Conclusion."

Unfortunately, analyzing the number of backers falls short when some projects have relatively low funding goals. A project looking to raise a few hundred dollars, for example, may only require a handful of backers to reach its target. As I was inclined to maintain parsimony and the model fit fairly well, I decided to not make this adjustment. In future analysis, the model can be enhanced with an adjustment for goal size.

Model Selection Criteria

I had three main dimensions along which I evaluated possible models:

- Fit – does the model fit the observed data well?
- Parsimony – is the model using as few parameters as possible, without jeopardizing fit?

- Story – do the parameter estimates make sense in the context of the dataset?

In order to evaluate fit, I calculated MAPE, BIC, and R^2 for each model. MAPE, short for mean absolute percentage error, is calculated by finding the percent error for each bucket in the histogram and then averaging all of those errors. BIC, short for Bayesian information criterion, offers a way to compare models that aren't nested and have varying numbers of parameters. Although all fit metrics were considered, I primarily relied on MAPE.

Benchmark Model

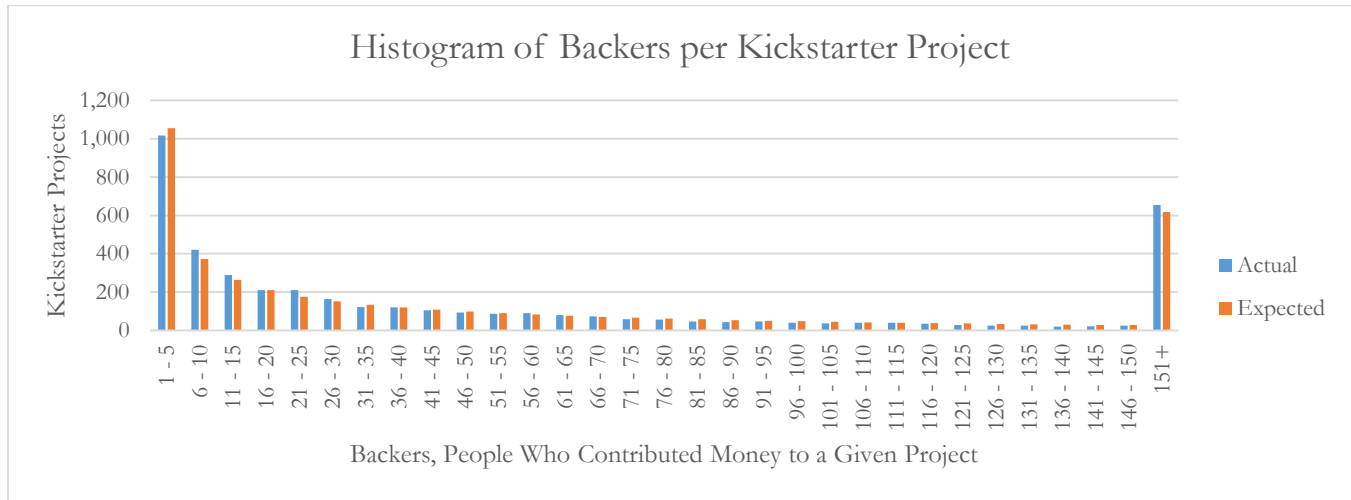
I found Empirical Bayes models, specifically the negative binomial distribution, to be appropriate for my use-case. Unlike regression, these models estimate a prior distribution based on the data, not exogenous factors. In order to provide intuition for how these models work, I will walk through the dynamics of the negative binomial model.

We start by analyzing the dynamics of one Kickstarter project and then scale up to the population level. Each project has a propensity (λ) to attract backers and that can be represented with a Poisson count process. The higher the propensity, the more backers a project will attract in a fixed period of time. Note that we will treat the duration of a project as a time interval of one, although the number of days a project remains open can vary.

Next, we introduce heterogeneity. While every Kickstarter project has an underlying propensity (λ), that propensity varies across projects. Some projects are endowed with better λ 's than others. Since λ can range from 0 to infinity, the gamma distribution is an appropriate choice to describe how propensities vary. The closed form model that results is the negative binomial distribution (NBD). I used this as the benchmark model. Note that since we excluded the projects that got 0 backers, this is technically called a shifted NBD.

Below is the resulting fit and histogram:

MAPE	14.08%
BIC	-23,876
R ²	99.38%



Alternative Models

Please refer to this summary as I walk through alternative models:

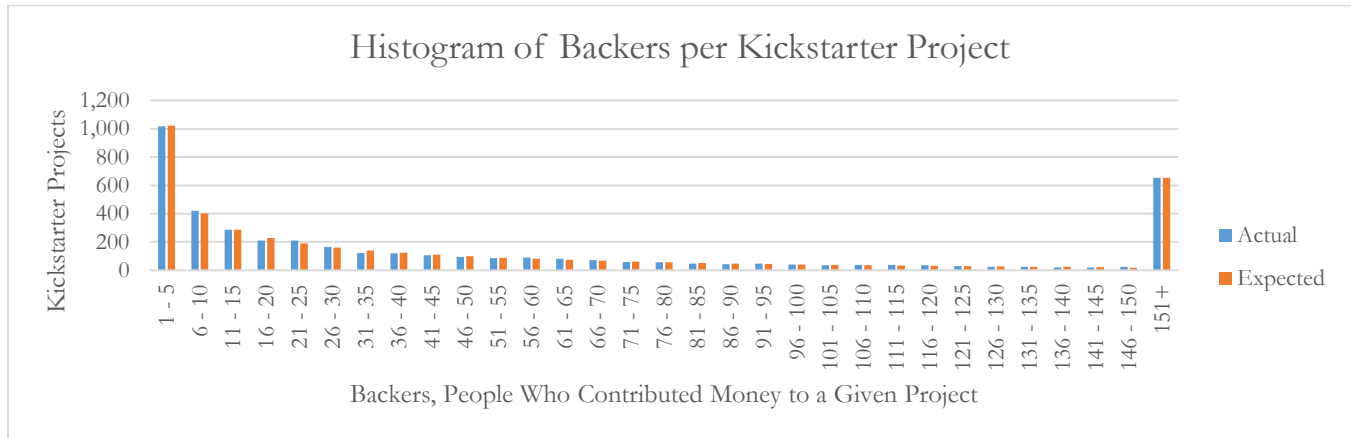
	MAPE	BIC	R ²	
Model 1	14.08%	-23,875.8	99.38%	Shifted NBD
Model 2	8.31%	-23,839.9	99.86%	Shifted NBD + Funding Window
Model 3	14.16%	-23,907.2	98.08%	Shifted NBD + Goal Size
Model 4	7.98%	-23,840.3	99.85%	Shifted NBD + Projects Created
Model 5	8.01%	-23,840.1	99.86%	Shifted NBD + Projects Backed
Model 6	7.36%	-23,820.9	99.86%	Latent-Class, Shifted NBD (2-segment)
Model 7	7.48%	-23,795.9	99.86%	Latent-Class, Shifted NBD (3-segment)
Model 8	8.01%	-23,831.7	99.86%	Shifted NBD + Projects Created + Projects Backed
Model 9	8.19%	-23,814.8	99.86%	Latent Class, Shifted NBD (2-segment) + Projects Created
Model 10	13.88%	-23,857.9	99.40%	2 Shifted NBDs - Data Split by Creator's Experience

Building off the benchmark model, I tried each covariate outlined in the “Dataset” section (Model 2-5). The covariates were scaled with a natural log because the percent change in the covariate is more relevant than the nominal change. Model 4 and 5 performed comparably and had reasonable parameters. I also tried creating latent classes (Model 6-7), which capture any existing sub-populations that behave differently. After evaluating the results of those models, I tried reasonable combinations (Model 8-10) but the additional complexity turned out to not be necessary.

Selected Model

Model 4, the Shifted NBD with “projects created” as a covariate was ultimately the best model.

First, the fit was remarkably strong, despite not having cleaned the dataset. Please see below the histogram for a visual comparison of the model’s results to the observed data:



The MAPE was 7.98%, meaning that the difference between the height of the blue and red bars above is 8% on average. The BIC was one of the best out of all of the models (closer to 0 is better). Since R^2 was high across all of the models, I did not give the slight differences in R^2 much consideration.

Second, the model was sufficiently parsimonious, utilizing only three parameters. The NBD uses two parameters and then there was the coefficient for the covariate. The three parameters are highlighted below:

0.485	r
0.012	α
1.991	β_{ProjCrtd}

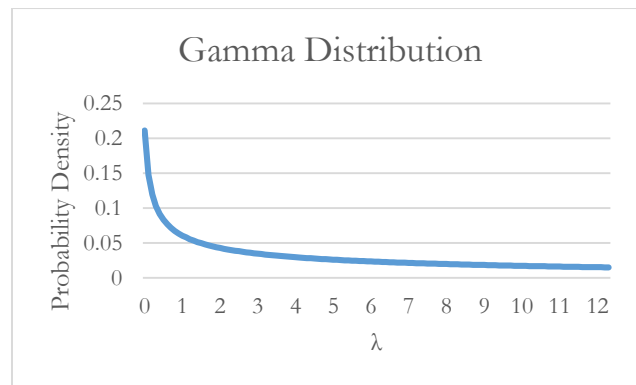
Third, the story the parameters told was a reasonable one. I will elaborate on this in the “Results” section.

Results

The Story

The selected model, (shifted NBD) + (projects created covariate), tells the following story:

The population of projects on Kickstarter is heterogeneous. This is supported by the fact that the r parameter is less than one (0.485). Some projects have high propensities to garner backers and others have low propensities, which is expected. The distribution of propensities is graphed below:



If the population were homogenous, there would instead be an internal mode.

The number of projects a creator has previously launched is highly associated with the number of backers a project will get. This is noted in the fact that the coefficient $\beta_{ProjCrtd}$ is positive and the model is substantially better (MAPE of 7.98% vs. 14.08% benchmark) when the covariate is included. The value of the coefficient can be interpreted as follows: when the project creator has launched one project previously (so this is his / her second), time feels stretched. Instead of having one period to accumulate as many backers as possible, it feels as though time is stretched and the project actually has 3.98 periods to accumulate backers. This table outlines how much time is stretched, based on how many projects the creator has launched.

<u>ProjCrted</u>	<u>Feels like time *</u>
1	1.00
2	3.98
3	8.91
4	15.80
5	24.64

Thus the story is simple – projects created is an important consideration in the success of a crowdfunding project. Of course, enhancements can be made to adjust and control for additional factors, but it is remarkable that such a simple story produces such a reasonable model.

Implications

This preliminary Empirical Bayes model provides regulators some amount of assurance that experienced project creators are the ones getting the crowd's support. As is the case with any new financial exchange platform, fraud is a concern. The more signs that the exchange is operating like a fully developed marketplace, the better.

The results are likely even more interesting to entrepreneurs, though. Knowing projects created is an important aspect of highly-backed campaigns can guide decision making. In order to be truly useful, though, we need to build a case for causation. There are three main causal forces that seem reasonable:

- Experience – creators that have launched multiple projects have learned from the experience and are now adept at executing a successful project
- Following – as creators launch projects, they accumulate a following, so gaining support becomes increasingly easy over time
- Person - the type of person that would launch multiple projects is often a skilled entrepreneur. In this case, the covariate lacks influence

Conclusion

Starting with a non-traditional definition of success, we have established that the success (many backers, rather than necessarily achieving a funding goal) of projects can be modeled with a reasonable fit, using only three parameters. The model tells a story of a heterogeneous population, affected by the number of projects a creator has launched. That covariate could encapsulate a number of forces as play – experience, following, or person.

In order to enhance our understanding of which force is strongest, we would need access to data on all projects launched by given creators. With that, we could generate natural experiments. For example, we could look at creators that launched multiple projects, spanning vastly different categories. If success is consistent across categories, then perhaps the “following” force is not at play.

As has been previously mentioned, access to the other half of data, the amount of money contributed by each backer, would enable a more holistic picture. At that point, we would have a mighty tool to yield. With both contribution amount and count data, customer life-time value calculations are possible, amongst other analysis. I am excited for that day to come when we have the data required for these next steps.

Bibliography

- Agrawal, Ajay K., Christian Catalini, and Avi Goldfarb. Some simple economics of crowdfunding. No. w19133. National Bureau of Economic Research, 2013.
- An, Jisun, Daniele Quercia, and Jon Crowcroft. "Recommending investors for crowdfunding projects." Proceedings of the 23rd international conference on World wide web. International World Wide Web Conferences Steering Committee, 2014.
- "Crowdfunding." Investopedia. N.p., n.d. Web. 14 Nov. 2014. <<http://www.investopedia.com/>>.
- Etter, Vincent, Matthias Grossglauser, and Patrick Thiran. "Launch hard or go home!: predicting the success of Kickstarter campaigns." Proceedings of the first ACM conference on Online social networks. ACM, 2013.
- Greenberg, Michael D., et al. "Crowdfunding support tools: predicting success & failure." CHI'13 Extended Abstracts on Human Factors in Computing Systems. ACM, 2013.
- Kuppuswamy, Venkat, and Barry L. Bayus. "Crowdfunding creative ideas: The dynamics of project backers in Kickstarter." UNC Kenan-Flagler Research Paper 2013-15 (2014).
- Lu, Chun-Ta; Xie, Sihong; Kong, Xiangnan; Yu, Philip S. "Inferring the Impacts of Social Media on Crowdfunding." WSDM, pages 573-582, 2014.
- Mollick, Ethan. "The Dynamics of Crowdfunding: An Exploratory Study." Journal of Business Venturing 29, pages 1-16, 2014.
- "Regulation Crowdfunding." United States. Securities and Exchange Commission. SEC Proposed Rules. N.p., 23 Oct. 2013. Web. 14 Nov. 2014. <www.sec.gov>.