Giving By the Numbers:
Peer Effects with Online Crowdfunding

A Thesis
Presented to
The Established Interdisciplinary Committee for Mathematics and Economics
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Claire Jellison

May 2020

Approved for the Division
(Mathematics and Economics)

Tristan Nighswander   Jerry Shurman

Acknowledgements

I want to thank a few people.

Contents

List of Tables

List of Figures

Abstract

This thesis investigates the size and nature of peer effects using data collected from the online crowdfunding platform, GoFundMe. On a campaign page, the previous five or ten donations are visible to potential donors. This allows for investigation into how the knowledge of these previous donations affects one's own decision to give as well as the size of the donation. I find that people are strongly influenced by the donations that are displayed to them prior to their own donation but the size of the effect is significantly different for men and women. Moreover, the distribution of the displayed donations is important in influencing how people decide to act.

Introduction

In microeconomics the decision of how much to buy is relatively straightforward. A person chooses how much to consume of each good by maximizing their utility, or overall satisfaction, subject to their budget constraint. Their utility, in turn, stems from a set of underlying preferences. These preferences are well-defined, transitive, complete, and constant with respect to time. Holding preferences constant and allowing relative prices to vary is how decision making is generally modeled. However, we know that people are not endowed with a set of immutable exogenous preferences and decisions are rarely made in such vacuum-like conditions.

Oftentimes consumption is visible which allows for social elements to enter into our decision calculations. With highly visible goods such as clothing, it's clear that people do not act independently; if this were the case, we wouldn't end up with constantly evolving trends and styles. We are constantly receiving cues about how people are spending their money, and it can influence how we choose to spend our own money. The choices of others matter to us when we're making our own decisions.

Not only do the choices of others matter, but what they would think of our own choices matters. One of the most well known sociological phenomena is conspicuous consumption. According to this theory, people spend disproportionately more money on goods that are more visible, like Rolex watches. That is because these goods convey status (Bagwell & Bernheim, 1996). Buying certain goods, can help to us craft an identity or at least a way that we want to be perceived. All of these social aspects, ultimately affect how we make decisions and they are highly context dependent, contingent on our knowledge of others' actions or perceptions.

This thesis examines how information about the actions of others drives our decisions within the context of charitable giving through analyzing data from the leader in online crowdfunding, GoFundMe. On the website the previous five or ten donations are visible at the time of donation, creating a natural field experiment where a potential donor is presented with a clear set of information points before making their own donation. This generates a two fold effect, people are prompted with the responses of others and they have the knowledge that their own donation will be visible unless they choose to make it anonymously. This thesis does not only attempt to confirm the existence of peer effects, but also seeks to better understand in a more general sense how people taken in multiple information points to make a decision.

Behavioral economics relies on the premise that while people do not always act the rational agent of consumer theory, they often deviate from it in predictable ways. Efforts in the field to understand how people act under different probability scenarios or when questions are posed with different framing, have demonstrated that people tend to do things systematically “wrong”.

This thesis follows in a similar vein of attempting to decipher generalizations about how people respond to information regarding the actions of others. In the first chapter, I review two relevant heuristics and biases generally referred to as anchoring and representativeness that feed into decision making and are closely tied to this study. I then review two distinct of models of giving that carry divergent implications for how people might respond to the donations of others. The second chapter details information about the choice of platform, method of data collection, and various attributes of the data. Finally, in the third chapter, I construct a variety of models ranging from linear regressions to algorithmic approaches to understand how people respond to the donations of others.

Chapter 1

Literature Review

1.1 Models of Giving

The larger part of US citizens give to charitable causes. In an average year, charitable donations now constitute more than 2% of GDP. Moreover, since 1968, the growth in charitable donations has greatly outpaced the growth of the S&P500. Individual donors account for over three quarters of the total number of gifts each year (List, 2011). This means that charitable donations are a non-negligible and increasingly substantial piece of the economy. Yet, beyond the magnitude of the charitable donation sector, the widespread nature of giving makes it worthy of attention especially within economics, a field dedicated to understanding the allocation of scarce resources. The challenge, for the discipline, lies in reconciling voluntary wealth transfers with the self-interested utility-maximizing agents that govern microeconomics. The basic question is what makes people compelled to give?

Models to answer this question generally fall into two categories. A critical point in the formulation of the two classes of models centers on whether the benefit from donating is primarily public or private. In the case of a public benefit both the donor and other individuals may benefit. For example, when donating to public radio the donor and others get to listen to the radio station as a result. In contrast, a private benefit is exclusively enjoyed by the donor. This encompasses the all the good feelings or abatement of negative feelings the donor might experience from giving. This distinction in the nature of the benefit proves important as it generates divergent implications for how people respond to changes in income or in the price of giving. Vesterlund (2006) notes that if we consider the benefit to be primarily private, then individuals will donate according to how much they value the product of a nonprofit or charity. Conversely, if the benefit is primarily public then someone else's donation provides the same benefits as one's own donation. Given that one can receive the same utility without incurring any cost, people are strongly incentivized to free-ride off of the donations of others. Consequently, when the donations of others increase in size, we would expect an individual's donations to decrease in response. Alternatively, if the benefit obtained by donating is primarily private then changes in the size of others' donations should have no effect on one's own donations. (Vesterlund, 2006).

1.1.1 Classical Model of Giving

The classical model focuses on the public benefit and assumes that the individuals derive utility from the nonprofit's output and their private consumption of the output, treating the contributions of others as given. In this model, each individual has little incentive to give and would be inclined to free-ride off of others' spending. This, in turn, leads to the possibility of complete crowding out (Vesterlund, 2006). In this situation, if the the government increases spending or funding to a certain service it could trigger a reduction of equal magnitude of private sector funding to support the same service. This is because the desired level of output can now be achieved without the contribution of the individual. Both of these implications, complete crowding out and extensive free riding, seem overly pessimistic and offer a poor depiction of what occurs in practice. Indeed, the classical model appears irreconcilable with the empirical evidence on charitable giving (Glazer & Konrad, 2008). Alternative models, which incorporate various private benefits, have been proposed that include avenues to reduce the propensity to free-ride and don't lead to complete crowding out.

1.1.2 Private Models of Giving

Private models of giving allow donors to receive a personal benefit that stems from the act of donating. That is, part of the benefit derived from charitable giving comes from making the donation rather than from the good that it buys. The private model no longer leads us to anticipate dollar-for-dollar crowding out when others donate to a cause that we are interested in; there is an added benefit to making the donation yourself that can't be experienced from another's donation. Further, these models provide a framework in which the incentive to free-ride is substantially weakened for the same reason.

The nature of the personal benefit that a donor obtains has received considerable attention within the charitable donation literature resulting in a large swath of theories designed to account for this donor-specific benefit. Harbaugh suggests that giving may enter the donor's utility function through two distinct mechanisms, the "intrinsic benefit", the donor's own knowledge of what they have given, and the "prestige benefit", which comes from other people witnessing the donor's generosity (Harbaugh, 1998). In an empirical study of lawyers' donations to their law school, he finds evidence that the prestige benefits accounts for a significant portion of donations. However, Harbaugh notes that the prestige benefit and the intrinsic benefit may not enter directly into the donor's utility function but rather as an amount relative to the gifts made by others. Glazer and Konrad (2008) construct a model where donors contribute in order to signal their wealth as with the theory of conspicuous consumption. They note that charitable donations may have some advantages over private goods, because unlike some forms of conspicuous consumption they may be less prohibited by social norms and can be visible to peers that a person does not directly come into contact with.

At the more extreme end of the private models of giving, Tullock suggests that the donor is not directly interested in the well-being of others. Rather, charitable giving allows one to purchase a feeling of satisfaction derived from one's financial sacrifice. This satisfaction in part stems from the benefit to the donor's reputation. He notes that the benefit to reputation is rooted in others' perceptions of the charity's image rather than its true effectiveness. This means that the gift enhances one's reputation when the people who matter to the donor consider the cause to be a worthy one (Tullock, 1966).

The "intrinsic benefit" that Harbaugh comments on is commonly referred to as the "warm glow" in the literature. Models including the so-called "warm glow" have proven to be more consistent with empirical observations and can also help explain diminished free riding and incomplete crowding out (Andreoni, 1989). Furthermore, if donors only care about the provision of a public good, then a charity has no incentive to announce donations sequentially. On the other hand, with donor utility functions that allow for effects like "warm glow" and "snob appeal", then the charity may benefit through sequentially announcing donations (Romano & Yildirim, 2001). Essentially, the announcement of donations converts the simultaneous game into a sequential game, if agents had standard utility functions, as in the classical, seeing others donations would likely reduce the size of their own, but with utility functions that are modified to include a private benefit the announcement can be advantageous. Through allowing individual's donations to be less responsive to the donations of others, these models ease the extreme consequences of the classical model.

1.1.3 Signals of Quality

A common fundraising strategy based on the private benefit model of charitable given involves soliciting "leadership gifts" to help inspire future contributions. These gifts are generally very large and occur at the onset of a fundraiser. Many hypothesize that this type of gift provides a signal to other donors about the quality of the charity that will encourage others to also contribute (Andreoni, 2006). In other words, if someone contributes a large sum of money other potential donors might assume that that person has obtained additional information about the quality of the nonprofit or charity and therefore be more inclined to give. Similarly, it has been shown that the total contributions to a cause tend to increase with the initial amount of seed money available for a campaign (List & Lucking-Reiley, 2002). Note that as a campaign approaches the goal amount, two effects emerge. On one hand, donors may be more compelled to free ride off of the donations of others since the fund is more likely to reach the goal even without their contribution. This effect is aligned with the idea that the benefit is primarily public; a donor is free to enjoy the output of the charity reaching its goal regardless of whether or not they played a part in attaining it. The other effect is the "follow-the-leader" component, where a large anonymous gift sends a signal of quality as in the leadership model of Andreoni. From the discrete amounts of seed money, they use in their study it appears that the "follow-the-leader" effect entirely subsumes the propensity to free ride. However, with a wider range of seed amounts (they use 3 levels 10%, 33%, and 67%), the interaction between the two effects is still unknown. This theory can be test with nearly continuous amounts of seed money using the data collected from GFM.

1.1.4 Gender and Anonymity

Another important facet of charitable donation is the interaction of gender with generosity. In a dictator style game, where the dictator divides a sum of money between themselves and a recipient, it has been shown that men tend to receive less money than women and men are more likely to withhold the entire sum of money (Dufwenberg & Muren, 2006). Another experimental study involving a modified dictator game finds that when the price of altruism is cheap, men are more generous, but when it is expensive women act more altruistically, implying that men are more sensitive to price changes (Andreoni & Vesterlund, 2001). In a study of door to door solicitation, it was discovered that females received more money when a male answered the door (Landry, Lange, List, Price, & Rupp, 2010). Given the nature of the data utilized in my study, I am able to contribute to this literature by leveraging insights from Sisco and Weber (2019), outlined below.

1.2 Heuristics and Biases

The classical and the private models of giving yielded differing narratives for how we would expect one's own donations respond to the donations of others. In the classical model, individuals are likely to decrease their donations if the donations of others increase. In the private model, the situation is slightly complicated, in theory the donations of others should largely not effect on one's own donation since one is rewarded with "warm glow" or "prestige" .Yet, the key feature in both the accounts of giving is that they both assume a rational actor with well defined underlying preferences. From there, they model decisions through a transactional lens; x units of money for y units of prestige. However, the idea of rational agents preforming calculations based on orderly preferences has been steadily undermined since initial efforts by Kahneman and Tversky, who introduced many of the heuristics, or mental shortcuts, that people systematically make use of to guide their actions, especially in the event of uncertainty.

1.2.1 Anchoring

Among Kahneman and Tversky's heuristics is a mechanism termed as anchoring. It captures the idea that people base their own predictions of values off of other known values or statements, especially in circumstances of uncertainty. Essentially, when prompted to act or respond to a given question, people tend to use some initial value, even if it is entirely arbitrary, as a starting point to adjust to their final response. However, different initial "anchors" result in different final estimates which are biased in the direction of the initial value presumably due to insufficient adjustment (Tversky & Kahneman, 1974).

In one of the most well known demonstrations of this effect, Kahneman and Tversky asked subjects to estimate the percentage of African countries in the United Nations. Prior to their estimate, a wheel was spun in their presence with numbers between 0 and 100. After observing the number displayed on the wheel, participants were asked to provide their best estimate of the percent. When 10 was spun the median estimate was 25 percent and when 65 was spun the median estimate was 45 (Tversky & Kahneman, 1974). This shows the strength that arbitrary values have in influencing people's estimates of reality, even when the anchors are unrelated to the question at hand.

While in this example, there is clearly a true value for the statement that people are attempting to estimate, the anchoring effect can also manifest in scenarios that lack a factually true value. In charitable donation scenarios, people are generally free to donate any amount. Here, the uncertainty lies not in the estimation of some true value, but rather in determining a course of action (i.e. whether to donate and, if so, the amount to donate) when it is unclear how one should act.

In one experimental study of charitable giving, undergraduate students were asked to complete a survey indicating how much they would give to a needy child. Groups of students were presented with a no anchor, a low anchor, or a high anchor. Anchors were statements of how many euros the average Italian would donate, with the high anchor condition corresponding to a large sum of euros and the low anchor to relatively few euros. They found that students presented with a high anchor were more likely to donate a higher amount than those who were given a low anchor or no anchor (Rubaltelli, Hysenbelli, & Rumiati, 2013). This study demonstrates how anchoring not only influences estimates of a true value, but can also help one determine the correct, or socially proper, way of acting.

The ambiguity of why people may anchor to a certain value poses some difficulty. Previously, in the wheel of fortune example, we saw how people were prone to latch on to arbitrary unrelated values in reaching their final answer. However, in the charitable donation scenario, another force is at play since the average amount given provides information about the actions of others in an identical situation. Although both anchoring and social influence could result in a similar observed effect, we might expect that the mechanism behind how they influence people's decisions is quite distinct. Therefore, it seems helpful to distinguish between anchoring and peer effects, or the propensity to conformity. Anchoring seems to generally involve having one's estimate biased towards any arbitrary initial value, while conformity involves basing one's estimate in a way that conforms to knowledge of how other people have acted. Therefore, the propensity to conform seems to be a distinct form of anchoring, given that anchoring can also occur when the stimulus does not involve the actions of other people.

1.2.2 Representativeness

Under the heuristic know as representativeness, Kahneman and Tversky detail people's insensitivity to sample sizes and erroneous conceptions of randomness. Citing several experimental studies, they claim that " ... intuitive judgements are dominated by the sample proportion and are essentially unaffected by the size of the sample, which plays a crucial role in the determination fo the actual posterior odds". Furthermore, they write that "People expect that a sequence of events generated by a random process will represent the essential characteristics of that process even when the sequence is short". They cite an example of this phenomena — in estimation of the likelihood of a sequence of coin flips H-T-H-T-T-H seems initially more likely than H-H-H-T-T-T. "...people expect that the essential characteristics of the process will be represented, not only globally in the entire sequence, but also locally in each of its parts." This "law of small numbers" is not limited to statistically naive subjects, but also manifests among professional researchers (Tversky & Kahneman, 1974).

This is relevant to the question at hand because people are not presented with a single value as in most experimental studies of anchoring, but rather see a distribution of past donations. If Kahneman and Tversky's claims about representativeness hold, then this implies that people view the displayed donations as more indicative of the overall pattern of donations to that campaign then they really are. However, certain distributions may appear more representative and therefore exert stronger anchoring effects than others.

1.3 Similar Work

The most closely related paper estimates the size of peer effects using a data set from two websites that people used to raise money for charity by running the 2010 London marathon (Smith, Windmeijer, & Wright, 2014). This paper looks at how very large donations, very small donations, the mode of past donations, and the mean of past donations affects subsequent donations. They find strong evidence for the presence of peer effects utilizing indicator variables for high and low donations as well as a linear-in-means model. Ultimately, they believe that donors give what they "think that they personally are expected to give where the distribution of the donations of their peers (along with other factors, such as income and specific cause) feed into the formation of that expectation". The difficulty is that because of the nature of their data, the mean of past donations is not really a consistent entity. As more donations arrive, each donation has less leverage in pulling the mean in either direction. They cite this as a finding — the later donations occur, the less influential they are. The issue is that there is an implicit assumption through using the cumulative mean that the sample size does not matter; the mean is treated the same by people regardless of how many donations there are. However, it seems dubious that people would scroll through all past donations, examine the distribution of past values, and act accordingly. The problem then, is that with their platforms, there is no clear indication of which donations people actually see before donating.

This GFM dataset is well suited for studying anchoring and peer effects because the past 5 or 10 donations are always displayed and the user must make an extra effort to see more donations. Furthermore, there are many facets of the site that allow for interesting comparisons. For one, the categories allow one to study, roughly, whether the effect varies in size thematically. Secondly, one can make inferences about anonymity, gender by coding first names, and familiarity by proxying through looking at the varying sizes of campaigns. Thirdly, the immense size of the dataset makes results more reliable and allows for more robustness checks.

A recently published paper (Sisco & Weber, 2019) appears to be the first to analyze a portion of the massive amount of data available on GFM. This study includes internet data from 9,264 campaigns to test several psychological hypotheses regarding the origins of altruism, specifically sexual selection and kin selection. According to the theory of sexual selection, certain traits may be selected for not because they increase one's likelihood of survival but rather because they serve in attracting mates thereby increasing reproductive prospects (Darwin, 1871). Alternatively, with kin selection, traits may be selected for that do not directly increase an individual's likelihood of survival, but that increase another's chance of survival who is likely to have the same genes (Stewart-Williams, 2015).Leveraging donor names to serve as a proxy for gender, Sisco and Weber find that women give with greater frequency while men donate less frequently but in larger sums. They find the larger sums donated by men are not enough to make up for the discrepancy in frequency of donations, as women's donations constitute a larger portion of the total (sum of all donations across all campaigns in their sample). However, they fail to find significant evidence that the campaign recipient gender affects the total amount raised. Furthermore, they find evidence that if the proportion of visible females on screen increases, female donors give less while male donors give more. They attribute this finding to the theory of sexual selection playing a critical role. Alongside gender, the authors also attempt to measure effects stemming from familial ties. They find that the average donation amount is significantly greater when a donor shares the same last name with the donation receiver. This could be attributable in part to the true kin effect they are trying to study but could also note it could be related to the idea that one empathizes more with someone of the same last name regardless of familial connections. In relation to anchoring, they find that the mean visible donation on screen at the time of giving was a significant and positive indicator of the value of donations of both men and women, with the effect being stronger in men. The empirical data analysis provides support for the presence of anchoring in charitable donation. It also leveraged to support the idea that kin selection and sexual selection play an instrumental role in altruism.

This project builds off of the work of Sisco in some minor ways (analysis of anonmity, category of campaign, etc), but most notably attempts to contribute to understanding *how* people anchor rather than if they do. To tell if people anchor the mean of past donations is a suitable candidate, but for understanding the mechanism behind anchoring it is not really sufficient. One disadvantage to using the mean of displayed donations as a hypothetical anchor value is that it assumes that people weigh each of the donations equally regardless of their relative magnitude. Consider three different possible sequences of displayed donations "50,50,50,50,50", "30, 40, 50, 60, 70", and "10,10,10,100,120". Although they all share the same mean, it's difficult to believe that people would respond in the same way to each of the sequences. However, understanding anchoring or peer effects with multiple reference points is not very well understood. In particular, it is unclear if people's donations will be more influenced large donations, very small donations, or donations that fall somewhere in the middle. However, on the website, a huge number of potential sequences are displayed allowing for investigation into how people respond to a multitude of reference points.

Chapter 2

Data Collection

2.1 GoFundMe

My main criteria when searching for potential crowdfunding platforms was evidence that they provided a clear anchor people would see before making a donation, there was a timestamp on each donation, and the site allowed scraping. The companies that I considered include GoFundMe, Crowdrise, Kickstarter, Indiegogo, Patreon, Fundly, and Kazoo. Kickstarter, Indiegogo, and Patreon were quickly ruled out because they focus strongly on a tiered donating approach, with several specified levels at which someone may choose to donate. Additionally, the donor may receive products or certain benefits based on their donation level. Therefore, they are out of scope for a project primarily focused on anchoring within the context of charitable giving. On the other hand, GoFundMe, Crowdrise, Fundly, and Kazoo all appeared to be good potential candidates.

GoFundMe provides five lags of previous donations to the viewer that could serve as an anchor as well as a record of all previous donations. Crowdrise, which is owned by GoFundMe, has ten lags of displayed donation as well as a complete donation history. Kazoo displays four lags and has complete donation history. Lastly, Fundly displays the ten highest donations and full donation history is available. All these sites were viable since they provide an anchor through the previously displayed lags and have complete time series data. I chose GoFundMe because it's the largest of all the sites considered and allows spiders to scrape the majority of it's website as indicated in the website's robots.txt [1].

GoFundMe claims to be the largest crowdfunding platform. Since its conception

[1] Visible at https://www.gofundme.com/robots.txt, which specifies to web crawlers what actions they are allowed to preform on the site

in 2010, the site had amassed over five billion dollars from 50 million people through 2017, when fundraising totals were last announced (Monroe, 2019). It's made headlines on numerous occasions, frequently appearing in the wake of national tragedies or alongside political movements Initially, GFM charged a five percent commission, but since 2017 has waived that fee for fundraisers in the US, although there is still a payment processing fee (Heller, 2019). There are 18 categories on GoFundMe including medical, memorial, emergency, non-profit, education, animals, business, community, competition, creative, event, faith, family, newlywed, sports, travel, volunteer, and wishes. The category is designated by the campaign organizer. Medical expenses are the most common, amounting to about a third of all campaigns(Monroe, 2019).

2.2 Web Scraper

I built the GoFundMe web scraper in python using the scrapy package, an open source web crawling framework designed to move through web pages by following specified links in the html source code and extracting data. Given that GoFundMe designates the order in which to display campaigns by some unknown algorithm, one of the initial questions was how to select data from the website at random. Sampling at random from all possible campaigns would ensure that the data was not biased by the characteristics that the website's algorithm selects for and promotes.

In order to counter the potential bias of the website's internal algorithm for ordering campaigns, I initially built one scraper that crawls through the pages of campaigns in a certain category collecting all the displayed links to each campaign page. However, after 1,000 unique campaigns in a given category, no more would be displayed by the site. Given the size of the website, this was clearly a very small subset of all campaigns. My impression is that campaigns that are selected to appear under the category labels, in general, have lots of donation momentum, high goal amounts, and polished profiles. In contrast, through searching certain keywords rather than just following category links, the disparity is evident; many of these campaigns are far from funding goals and have received little to no attention. Therefore, sampling from the category pages would be highly subject to the logic underlying the site.

Using the search feature provided a way to bypass the website's campaign ordering logic. The search feature on the website allows a viewer to input certain keywords and then pulls up relevant campaigns that use that word. The issue once again, was that while a certain term might say that it yielded millions of results, only around 1,000 would actually be displayed rather than the full set of results.

The key to resolving this issue was by using zip codes. Within the search feature, it is possible to specify the country as the US and also include a specific zip code. This allowed me to build a scraper that searches a given zip code on the site and then scrapes all the campaign urls for that zip code. This works because every campaign on the site is necessarily tied to a zip code and thus the full list of zip codes partitions all US campaigns.

I downloaded a full list of US zip codes[2]. I randomly partitioned all the zip codes. I passed this list of zip codes from the first partition to a scraper with three parsing methods. It first completes the search for each zip code from the homepage. These links for each zip code are sent to the first parsing method which pulls all the urls for campaigns attached to that zip code. These links go to the second parsing method which records several campaign specific traits and generates links to the donation history data. These links are sent to the third parsing method which calls itself recursively until all past donations have been recorded. An outline of the scraper can be seen in Figure 2.1 [3].

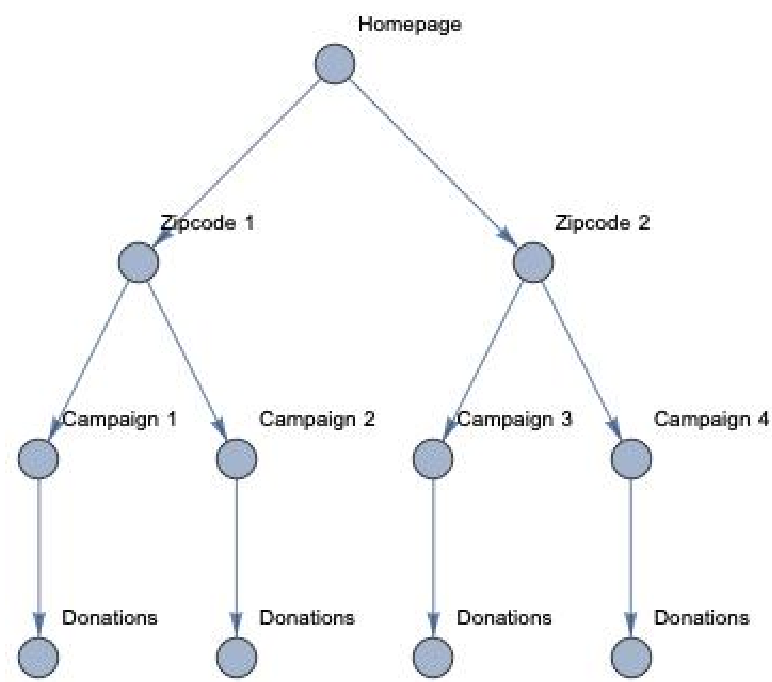

Figure 2.1: Schema of scraper

[2]https://www.unitedstateszipcodes.org/

[3]Source code for the scraper can be found on my github account, https://github.com/zumiko/thesis

2.3 Data Description

The campaign zip code table provides the zip code for every campaign scrapped and a unique key for identifying each campaign. The time history data for each campaign includes the campaign key, the amount donated, the time of the donation, the donation identification number, whether the donation was anonymous, the self entered name of the donor, as well as the fields “isoffline”, “profileurl”, and “verified”. The campaign attributes data table includes the key for identifying the campaign, the title of the campaign, the category of the campaign, the date that the campaign was created, whether the campaign has been terminated, the campaign organizer, the campaign description if provided, whether the campaign is a registered tax deductible non-profit, the amount raised, and the goal amount.

A total of 21,635 campaigns were collected, with a total of 849,764 individual donations. The mean donation amount was around 115 and the median was 50. For anonymous donors, the mean donation amount is significantly higher at 145. Approximately, 23 percent of donations were made anonymously. Of all campaigns, around 7.3 percent were tax deductible nonprofits.

In order to gain insights into how gender interacts with charitable giving, first names were assigned a gender, male or female, based on their likelihood of being that gender according to historical data using the gender package in R (Mullen, 2019). Some observations could not be classified and are listed as unknown. These are due to names with ambiguous gender, anonymous donations, couples donating,(i.e. “Mr. and Mrs. Smith”), donors using their relationship to the recipient (i.e. “Your Grandma”), or other variants. Obviously, this method is not ideal and thus any interpretation of the results relating to gender should be regarded with ample suspicion.

It is likely that GFM usage may not be uniform across the US and therefore the data may be more representative of certain areas of the country based on where the website is more popular. However, the website is used quite broadly across the US although the demographics of who uses the site is not well known. Additionally the origin of where a donation was made from cannot be determined from the data only the location of where the campaign is located. Below, we can see the geographical configuration of the data sample in Figure 2.2.

A significant drawback of the data, is that the exact date that the website switched from displaying the previous ten lags to the current layout where the previous five lags are displayed is not exactly known. Given the previous study Sisco and Weber

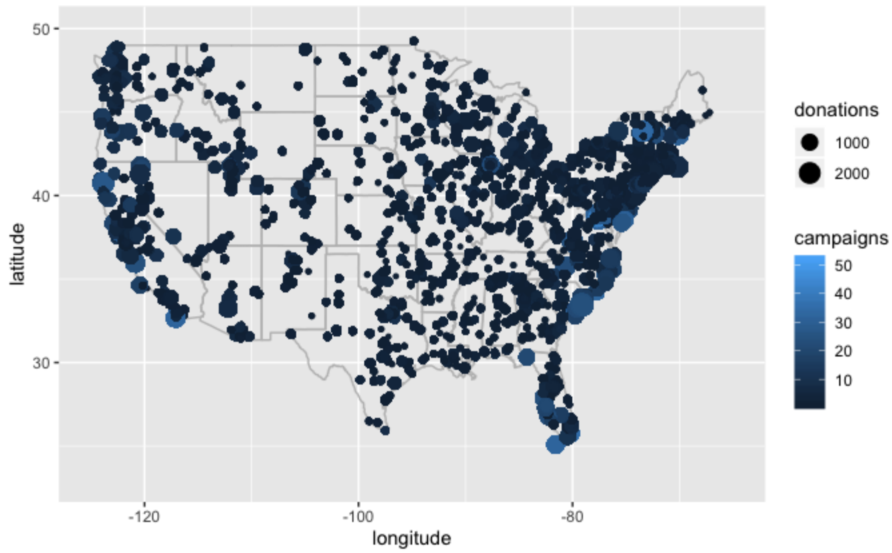

Figure 2.2: Map of data on contiguous states

(2019) who collected data in June 2016, the layout change occurred sometime after that month so all donations previous to that month are coded as having ten lags. Similarly, since I began to work on this project, the site has displayed 5 lags so these donations can also be appropriately coded. However, a large portion of the donation data had to be discarded for lag analysis since the exact date is unknown. Checking screenshots of the website posted online could likely narrow the window of donations that cannot be coded.

A natural question to ask is how often campaign goals are met. Since few campaigns are ever terminated a cutoff for determining inactive campaigns was needed. I chose to look at all campaigns that had not received a donation within the past year and considered these campaigns to be inactive. In theory, these campaigns could still receive more donations but the chances are very slim given that a year has passed. For campaigns that met this criteria, a percent of goal met was calculated by dividing the amount raised by the goal amount. In Figure 2.4, a histogram of the inactive campaigns progress towards their goal is displayed. This is particularly relevant to the aforementioned work on seed money by List & Lucking-Reiley (2002) who found that increasing the amount of initial seed money resulted in more donations of greater size. Since the amount of seed money is necessarily greater for each subsequent donation, we would expect fewer campaigns to stagnate at levels closer to their goal amounts. Since the probability of donations increases with additional seed money, we would expect that the number of campaigns at each level of percent raised level to decrease which appears to be the case in the figure. Of course, it's possible that there are reasons for the distribution.

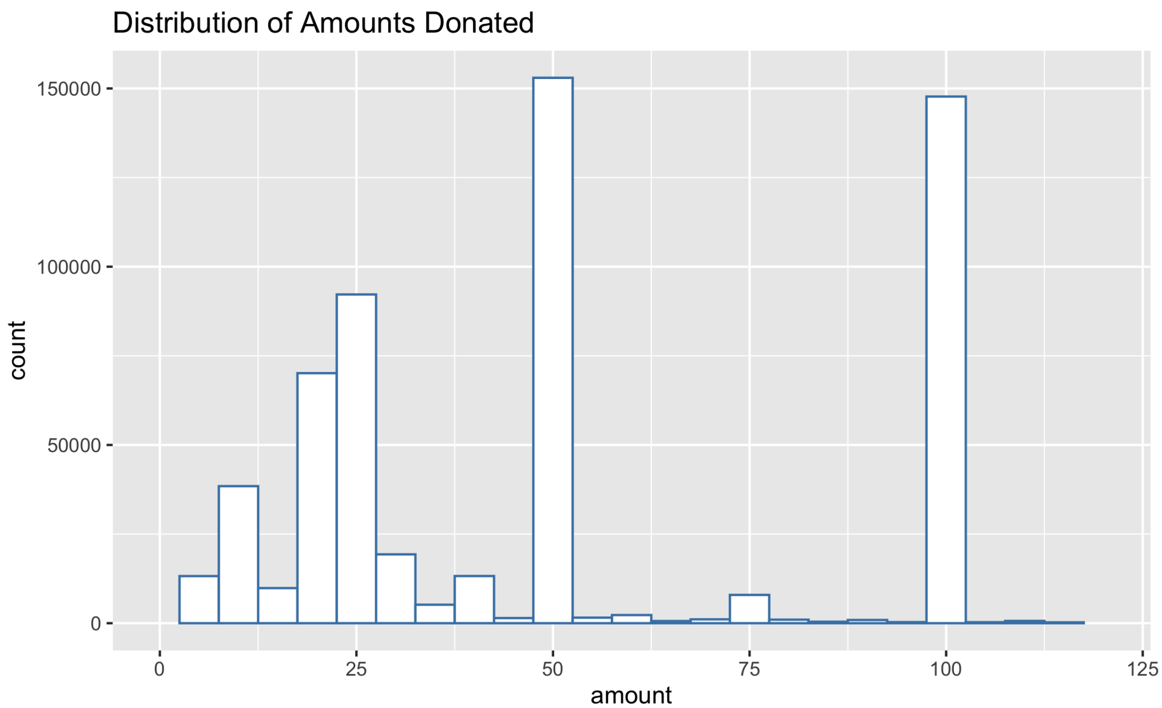

Figure 2.3: Distribution of donation amounts

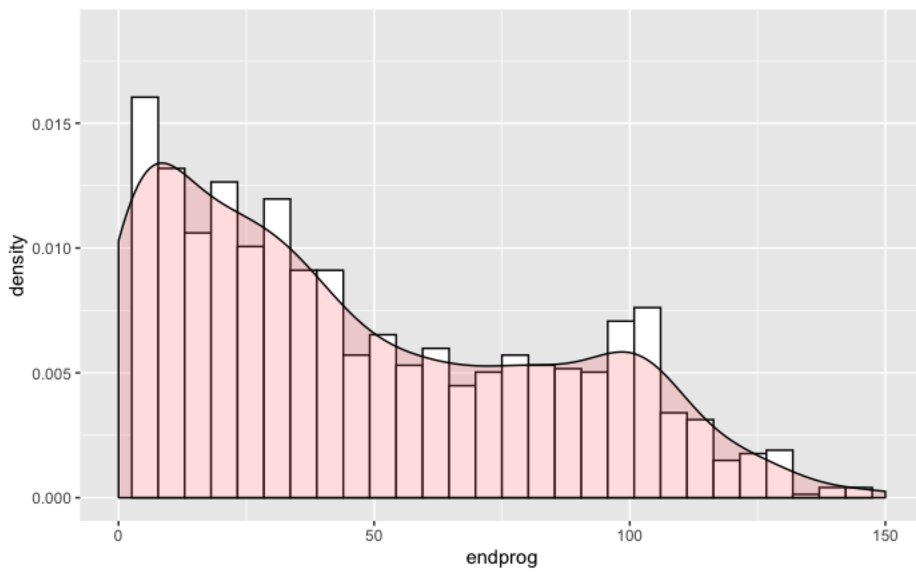

Figure 2.4: Distribution of percent funded

Chapter 3

Regression Models

3.1 OLS Models

Perhaps the most obvious choice for an aggregate value to test for the presence of anchoring is the mean of the displayed donations. Let m be the number of donations displayed on screen (this is necessary because the site initially displayed the previous 10 donations but later modified the layout to only include the previous 5 donations). This means that for y_j, the jth donation, the mean anchor value is calculated as $a_j = \frac{1}{m}\sum_{i=1}^{m} y_{j-i}$ where $j > m$. To test for whether this naive anchor showed any significance as well as determine whether the effect had any interaction with gender and anonymity, the several linear models were estimated [1].

From the Table 3.1, we see that the mean of displayed donations appears highly significant. For every one dollar increases the mean we anticipate between a 14 cent and 24 cent increase in the subsequent donation with all else equal. We also see that donations coded as female tend to be lower than male donations and more notably have a significant interaction with the anchor, demonstrating that women are less influenced by the values of the displayed donations. Anonymous donations, may be slightly larger but interestingly conform less to the mean of displayed donations than non-anonymous donations. The R^2 values range between .01 and .02, suggesting that these model have very low explanatory power, explaining only around 1-2% of the variance in the amount donated. Nevertheless, predictors have extremely low p values meaning that they do have an effect on the amount donated. Although, this may seem somewhat contradictory, it is not a surprising result given that there are

[1] All R source code for various models can be found on my github account, https://github.com/zumiko/thesis. Additionally, residual plots and tests for stationarity are included in the R files

	Model 1	Model 2	Model 3	Model 4
(Intercept)	75.07***	74.19***	98.19***	69.95***
	(2.78)	(2.89)	(2.30)	(3.41)
histavg	0.24***	0.24***	0.14***	0.26***
	(0.01)	(0.01)	(0.01)	(0.02)
lagtypeten	0.84	−0.44	−14.01***	5.62
	(3.48)	(3.49)	(2.25)	(4.00)
histavg:lagtypeten	0.00	0.01	0.06***	−0.02
	(0.01)	(0.01)	(0.01)	(0.02)
is_anonymous		10.37*		
		(4.10)		
histavg:is_anonymous		−0.08***		
		(0.02)		
is_female			−24.33***	
			(2.24)	
histavg:is_female			−0.11***	
			(0.01)	
nonprofitTRUE				50.13***
				(9.71)
histavg:nonprofitTRUE				−0.02
				(0.02)
R^2	0.02	0.02	0.01	0.02
Adj. R^2	0.02	0.02	0.01	0.02
Num. obs.	166769	166769	116083	165200
RMSE	649.97	649.94	351.11	652.58

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.1: OLS regressions for mean of displayed donations

a host of omitted variables that would undoubtably exert considerable influence on how much a person might donate such as level of income and relation to the recipient. Unfortunately, it is difficult to compare this model to the similar work by because R^2 values are generally excluded from model summaries. An additional model that includes categories from the site and interactions with anchoring is included in the first appendix.

3.2 Size and Order variant models

While the previous models are highly indicative of the existence of anchoring, they do relatively little to explain how an anchor is formed. Rather, basing prediction off of the mean implies that all displayed values are weighted evenly, or taken into consideration independent of their size, name, gender, and order. In attempt to increase the explanatory power of our model and we'll now take into consideration the order and the size of the donations. My initial hypothesis was that the order of the five donations would be largely irrelevant. To test this hypothesis, I estimated a model that of the form $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + \beta_5 y_{t-5}$. Then I randomly permuted the lag values and ran the same regression. Then I tried permuting the lags again but with the first lag as a fixed point. The results are shown in Table 3.2.

The most notable observation is that including all five displayed donations significantly improved the explanatory power of the model. Nevertheless, the gains in explanatory power remain largely unaffected by random permutations. It seems plausible that the most recent donation has some influence due to its position on screen. However, the R^2 values are only very slightly lower for the two randomized permutations. The same procedure was done for the subset of the data that contained 10 visible donations. In this case, the original R^2 was 0.007 and after the first random permutation actually increased to 0.009. Therefore, order seems only minimally important if at all for the purpose of explaining future donations.

Given that order seems of minimal importance, we now turn our attention to size. Looking at the donations according to their relative size can help us to gain insight into whether relatively high or low donations exert more control over a future donation. Allow five new variables d_1, d_2, d_3, d_4, d_5 so that for the jth donation, d_1 would correspond to the smallest donation displayed and $d_1 \leq d_2 \leq d_3 \leq d_4 \leq d_5$. This permutation allows us to see which donations, relatively smaller or larger, are more important in determining the size of the subsequent donation. The results for

	Lags	Permutation 1	Permutation 2
(Intercept)	72.08***	77.18***	73.05***
	(1.67)	(1.70)	(1.71)
lag1	0.19***	−0.00	0.10***
	(0.00)	(0.00)	(0.00)
lag2	−0.03***	0.04***	0.02***
	(0.00)	(0.00)	(0.00)
lag3	0.13***	0.17***	0.14***
	(0.00)	(0.00)	(0.00)
lag4	−0.03***	0.02***	0.00
	(0.00)	(0.00)	(0.00)
lag5	0.00	−0.02***	−0.00
	(0.00)	(0.00)	(0.00)
R^2	0.16	0.15	0.16
Adj. R^2	0.16	0.15	0.16
Num. obs.	56959	56959	56959
RMSE	388.38	390.77	389.23

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.2: Permutation of order of lags

the linear regression with these new variables is displayed in Table 3.3[2]. The size of the coefficients on this model are significantly larger than in the previous order based regressions. In particular, the large coefficient for the smallest donation seems to indicate a tendency to not want to be the lowest donation on screen. Perhaps, the most confusing element is the negative coefficient on the second and third smallest displayed donation. This leads me to believe, that it is not only important to consider the relative magnitudes but also their distribution, or the spacing between them.

We might expect that five densely packed observations would influence a person differently than five very spread out donations or similarly that donations clustered on the lower end of the range could behave differently than donations clustered at the higher end of the range. I took a couple approaches to looking at the effect of the distribution of displayed donations. To more explicitly control for the distribution of values, let d_{ij} be the distance between the ith smallest and the jth smallest donation so that $d_{ij} = d_j - d_i$ for $j > i$. The results for the new model can be seen in Table 3.4. Now, the variable d_1 encodes the effect of taking all five donations and moving them higher or lower but keeping them spaced exactly as they were before. The other variables have a slightly less interpretable meaning; as an example, a change in d_{12}

[2]This model is just using the subset of data where five donations were present on screen

	All displayed	Min	Median	Max
(Intercept)	25.02^{***}	12.68^{***}	82.53^{***}	78.22^{***}
	(2.08)	(2.08)	(1.79)	(1.73)
d1	2.94^{***}	3.19^{***}		
	(0.06)	(0.04)		
d2	-0.40^{***}			
	(0.03)			
d3	-0.39^{***}		0.25^{***}	
	(0.02)		(0.01)	
d4	0.16^{***}			
	(0.01)			
d5	0.08^{***}			0.08^{***}
	(0.00)			(0.00)
R^2	0.14	0.08	0.03	0.08
Adj. R^2	0.14	0.08	0.03	0.08
Num. obs.	56959	56959	56959	56959
RMSE	392.67	405.98	417.82	407.09

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.3: Relative size of displayed donations regression

would mean increasing d_2, d_3, d_4 and d_5 the same amount since this keeps all other distance variables equal and d_1 does not change. Looking at the coefficient on d_1, suggests a tendency to donate slightly higher than the lowest donation on screen.