THE EFFECT OF PEER RATINGS ON NONPROFIT CONTRIBUTIONS: EVIDENCE FROM CHARITY NAVIGATOR

by

Tavio Aleksandrs Pela

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Applied Economics

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2024

ACKNOWLEDGEMENTS

First, I would like to issue a heartfelt thank you to my defense committee, whose invaluable feedback and expertise pushed and sustained me throughout my research. Sincere thanks to the chair of my committee, Dr. Andrew Hill, for his continuous patience and active technical and theoretical support through my time of study. His cheerful optimism and clearheadedness helped me navigate numerous challenges and maintain focus on the proper next step. Thank you to Dr. Carly Urban for both unwavering frankness and friendliness in feedback, as well as alerting me to additional strands of literature I could bring to bear on my topic of interest. Last, but certainly not least, an emphatic thank you to Dr. Christiana Stoddard for early feedback that helped solidify the direction of my research and for ongoing aid in refining both my model and my writing.

Significant contributions to this work were also made possible by other professors, both within and beyond Montana State University. I am indebted to Dr. Richard Patterson from Brigham Young University for suggesting the firm-peer pair stacked data structure that underlies this paper's primary methodology, and I was greatly aided in my ability to perform the webscraping necessary to acquire key data for this research by coursework from Dr. Nick Hagerty and Dr. Wendy Stock.

I am also grateful for the moral support provided by my graduating cohort members and my friends and family. In particular, special thanks to my partner Sarah, who has seen me through every roadblock with encouragement and understanding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

ABSTRACT

Nonprofit rating organizations publish third-party assessments of nonprofit organizations for current and prospective donors. Using charity-level yearly financial and ratings data for organizations rated by Charity Navigator, a prominent nonprofit rating organization, I employ a regression discontinuity design to investigate whether a charity's total contributions are impacted by changes in the ratings of its competitors. I find a negative relationship between current-period peer rating and current-period contributions, which is consistent with peer ratings being used to inform donation decisions between comparable organizations. However, difficulty substantiating a key identifying assumption of the RDD raises doubts that these findings identify a causal relationship.

1
INTRODUCTION

As of 2020, there are over 1.8 million US nonprofits registered with the IRS [7], but not all charities are created equal. Services addressing similar problems can vary significantly in cost and impact. For instance, training a seeing eye dog for a blind person costs around $40,000, while reversing severe visual impairment with cataract surgery only costs about $1,000 [8]. Nonprofits may also be ineffective or dishonest. In the absence of informed giving, competition between charities does not guarantee quality, instead increasing the percentage of charity scams that fundraise for personal enrichment and provide no public good [26].

Numerous informational challenges face donors navigating the charitable market. Information asymmetry arises because donors often do not directly receive the final output produced by charities. The social benefit of said output can also be difficult to quantify, especially for more abstract final products like those provided by libraries and museums [46]. The observed demand for philanthropic goods and services comes from the donors who finance them rather than their recipients. The lack of profits or product pricing for many charities further makes it difficult to quantify the value added to recipients of their services [32]. Moreover, donors who are primarily motivated by the subjective psychological benefit derived from giving, social pressures, or desires to alleviate guilt may also be less motivated to seek information and less sensitive to indicators of quality than donors motivated by the overall level of charitable services provided [10].

Nonprofit rating agencies, also referred to as charity "watchdogs," are one mechanism

for providing third-party accountability and information on nonprofit quality to prospective donors. Watchdogs do this by publishing assessments of nonprofits in forms such as 'Pass/Fail' ratings or '0-100' scores. Notable rating agencies include the Better Business Bureau's Wise Giving Alliance, CharityWatch (formerly known as the American Institute for Philanthropy), and Charity Navigator—the largest charity evaluator in the United States [24]. However, usage data alone is insufficient to indicate whether third-party charity ratings affect donor behavior and play an informational/disciplinary role in the market for charitable contributions. About two thirds of US donors do not use nonprofit rating organizations, though twenty-one percent of donors always or usually use ratings to help evaluate nonprofits, [6]. Awareness of charity watchdogs is higher among younger generations, indicating that usage may grow over time. Estimates put the number of donors using watchdogs in the tens of millions—a non-trivial quantity, especially since users are likely to be donors who give large sums of money or are engaged in advocacy [14].

Numerous scholars have empirically investigated whether the rating a charity receives influences the contributions it receives, and they generally find a positive relationship between own-rating and own-contributions for the ratings systems of various watchdogs [12, 13, 20, 35, 47][1]. However, literature on the relationship between the ratings a charity's near-competitors receive and its own contributions is nearly nonexistent. Using data from Charity Navigator's website and the National Center for Charitable Statistics, I construct a charity-level panel of yearly ratings and financial data between 2000 and 2010 for 448 charities organized into

[1]Authors also typically find heterogeneous effects depending on rating type and size of charity, which will be discussed further in the Background chapter of this work

117 peer groups. I estimate the effect of one of a charity's peers receiving a higher star rating from Charity Navigator on the charity's contributions using a regression discontinuity design. The results provide evidence that rises in the star rating of a charity's peers to the three- or four-star level are negatively related to that charity's contributions in the same period. However, these results may not be indicative of a causal relationship due to concerns that the identifying assumptions of the RDD do not hold.

4
BACKGROUND

Charity Rating Agencies

The first nonprofit rating agency in the US was the National Charities Information Bureau, founded in 1918, which merged with the Better Business Bureau's Philanthropic Advisory Service in 2001 to form the BBB Wise Giving Alliance [40]. This organization rates nonprofits on a pass/fail basis according to 20 standards that incorporate information such as governance, effectiveness, and finances [2]. As additional rating organizations like the American Institute for Philanthropy (AIP) and Charity Navigator emerged, the methodologies and criteria used varied significantly among the various rating and ranking organizations, and many were criticized for inadequately incorporating program effectiveness [4]. Evaluations from AIP and Charity Navigator were highly correlated, but seemingly unrelated to the number of BBB standards met, possibly because they focused on financial assessments while the BBB had a more balanced system [37]. Over time, these organizations have continued to refine their methods. For example, Charity Navigator initially relied on financial data alone for its ratings, but in 2011 expanded its system to encompass two key dimensions, "Financial Health" and "Accountability and Transparency," and it moved to its current four-dimensional system comprised of Finance & Accountability, Impact & Results, Leadership & Adaptability and Culture & Community in 2021 [1].

This paper focuses on ratings from Charity Navigator. Founded in 2001, Charity Navigator rates 501(c)(3) tax-exempt organizations using information from their annual

IRS Form 990 filings and their websites.[1] Rated nonprofits are assigned a numerical score in various key categories that is combined into an overall numerical score on a 100-point scale, which is then used to give the organization a rating of zero to four stars.[2] Both star ratings and score ratings for organizations are displayed on Charity Navigator's site [1].

To receive a rating from Charity Navigator, organizations must have total revenues of above $1 million in the most recent fiscal year, with at least $500,000 in public support, and public support (which includes direct contributions, indirect contributions, and government grants) must make up at least 40 percent of total revenue for two consecutive years [15]. The site does not review charities that receive most of their funding from government grants or program/service fees [20].

During the years rating systems were becoming more prominent, organizations' reported expenses shifted in ways favorable to ratings metrics, but this may be a result of organizations changing how they report costs—"playing to the test"—rather than a result of real changes in their spending behavior [38]. Indeed, there is some evidence that organizations highly rated by Charity Navigator disproportionately underreport their fundraising expenses and

[1]Form 990 is a tax form all 501(c)(3) tax-exempt organizations must file annually that includes data on income and expenditures and other information used to assess whether the nonprofit meets federal requirements for its tax-exempt status. These forms are made publicly accessible at numerous sites, generally after a 12–18-month delay following the end of an organization's fiscal period [45].

[2]Scores displayed on Charity Navigator's website range from 30 to just under 100 in the years 2002 to 2010. During this period, Chairty Navigator actually calculated score on a scale of 0 to 70, and later linearly transformed all these past ratings by adding 30 when it transitioned to a 0-to-100-point scale in 2011. Thus, when examining current data for that time period, cutoffs for different star ratings are thus at 90 (four stars), 80 (three stars), 70 (two stars), and 55 (one star).

overstate the fraction of expenses they allocate to programs rather than overhead [27]. However, nonprofits also make real changes to spending when rated. Chhaochharia and Ghosh (2008) find evidence that charities react to lower ratings by the American Institute of Philanthropy by increasing their fundraising expenditures, and Grant (2017) finds that organizations rated by Charity Navigator typically decrease fundraising expenses, perhaps disproportionately enough to reduce charitable contributions beyond what they could earn without damaging ratings.

Impact of Charity Ratings on Contributions

The relationship between a charity's ratings and the contributions it receives has been studied with data from numerous rating organizations. Most authors find evidence of a positive relationship, albeit with caveats. Using data from charities in New York with an emphasis on New York City between 2000 and 2001, Sloan (2009) finds that BBB Wise Giving Alliance "pass" ratings have a statistically significant effect on the contributions received, but that "did not pass" ratings are not significant. Similarly, Chen (2009) finds a significant effect of meeting BBB standards on the amount of public giving to NY charities with data from the years 2005 and 2006. These studies both employ OLS regressions. Chhaochharia and Ghosh (2008) use ratings data from the American Institute of Philanthropy from 1999-2004 and find a positive relationship between current-year contributions and last-year rating, even controlling for time, state, and charity fixed effects.

A few major studies have investigated Charity Navigator's ratings specifically. Earlier

studies with less credibly causal designs find mixed results. Gordon et al. (2009) use a random sample of 405 rated charities and find that contributions in the current period are related to both the change in star-rating and the star-rating received in the prior period. They include a control for year fixed effects, but not organization-level fixed effects. Szper and Prakash (2011) use OLS regressions with data for the period 2004–2007 on all rated charities in Washington State, finding no association between changes in rating and changes in either primary revenue or contributions.

Taking advantage of the fact that an organization's star-rating is determined at specific thresholds of score-rating, Yörük (2016) employs a regression discontinuity design to test the effect of a one-star increase in Charity Navigator ratings on contributions. Using a sample of all charities rated by Charity Navigator from 2007 to 2010, he finds that while there is no significant effect for the full sample, the ratings have a considerable impact on the amount of charitable donations received by relatively small charities.[3]

Grant and Potoski (2015) is the only study to examine peer rating effects. They use yearly ratings and financial data from 1993-2008 for 3,474 nonprofits at some point rated by Charity Navigator that operate on a state level or smaller to estimate the effect of peer ratings on charity contributions. They define peer groups as "those nonprofits in the same state and with the same category of policy mission" and peer ratings as the average star-rating within these peer groups. The researchers impute Charity Navigator ratings for years before a charity was actually rated and find that the effect of peer ratings when the ratings

[3]Yörük (2016) defines charities as "small" in a given year if they have lower fundraising, contributions, and assets than the medians of those values for the full sample.

of the nonprofit and its peers are both published is significantly greater than the effect of peer ratings when neither ratings are published using fixed effects regressions and two-stage least-squares regressions. They ultimately conclude that the positive relationship they detect between peer ratings and a charity's contributions appears to be a result of charities increasing their fundraising as their peers are rated more highly.

Indeed, though peer ratings could have a positive effect on a charity's contributions if they contribute to creating a sort of "collective reputation", they could have a negative effect if they are used to make donation decisions between substitutable charities. Evidence that expenditure substitution can occur between charities is present in both experimental literature [18, 30, 33] and other empirical literature [11, 34, 41], though this latter research typically does not quantify the effect's magnitude.[4] Thus, I hypothesize that higher peer ratings could exert a negative pressure on contributions independent of the indirect positive effect they may have through encouraging higher fundraising.

[4]Under certain circumstances and between specific types of causes, there may be no charitable substitution. See Deryugina and Marx (2021) for a recent paper which fails to find evidence of charitable substitution after deadly tornados, suggesting that giving in response to new needs does not always trade off with giving to other causes.

9
DATA

Historical charity ratings data was obtained via a webscrape of Charity Navigator's website conducted between 10/11/23 and 11/14/23. For each historical rating, the data includes the date the rating was published, the fiscal year of the Form 990 financial data the rating was based on, the overall numerical score the charity received, and the number of stars it was given.[1] Data was available for 3,893 unique organizations rated at some point between the years 2002 and 2010. Yörük (2016) reports that 5,400 organizations have Charity Navigator ratings data between 2007 and 2010. The likeliest explanation for this discrepancy in amount of organizations is that my data collection methodology is unable to collect ratings data for organizations which have ceased to exist and thus no longer have data on Charity Navigator's site.[2]

Financial, industry classification, and location data was obtained from the National Center for Charitable Statistics (NCCS), the national clearinghouse of data on the nonprofit sector in the United States.[3] Industry classification data consists of a charity's classification under the National Taxonomy for Exempt Entities – Core Codes (NTEE-CC), which is

[1]On 10/19/23, Charity Navigator updated the historical ratings section of its website in a way that removed the ability to see the overall score and Form 990 FYE information, so the Internet Archive's Wayback Machine was used to scrape this data for the remaining charities (https://archive.org/web/). The Internet Archive is a nonprofit organization that serves as a digital library storing present and past copies of web pages [3].

[2]Many organizations rated by Charity Navigator and present on their website do not have any historical ratings data listed on the site, but this only means they were rated fairly recently (M. Viola, personal communication, January 26, 2024). Nonprofits also go under relatively frequently, with one study from the National Center on Charitable Statistics estimating that approximately 30% fail to exist after 10 years [17].

[3]Specifically, I use the NCCS Core Files, available from the Urban Institute at: https://nccs-data.urban.org/data.php?ds=core.

discussed further in Section 4.2. State and county FIPS codes are merged with data from the Bureau of Labor Statistics to identify organizations in the same Metropolitan Statistical Area.[4] NCCS data is available for all but 11 of the rated organizations and is not available in every single year for each organization. Thus, the full sample consists of an unbalanced panel of annual ratings and financial data for 3,882 organizations from the fiscal reporting periods ending in 2000 to 2010. Summary statistics are detailed in Table 3.1.

Table 3.1: Charity Summary Statistics (Full Sample)

	N	Mean	SD	Min	Max
Total Contributions	38653	139.86	529.93	0	33852.16
Score	25018	83.31	10.11	30	99.94
Stars	25018	2.85	0.99	0	4
Fundraising Expenses	38629	9.49	43.57	0	1685.95
End of Year Assets	38653	449.17	2311.22	-9.07	113972.9
Program Revenue	38632	31.71	443.88	-2.80	29239.64
Total Expenses	38651	172.36	839.60	0	61208.63
Fundraising/Expenses	38628	0.08	0.09	0	1
Δ Total Contributions	32012	3.57	272.63	-26780.34	18122.66
Δ Score	21163	0.18	7.61	-40.99	53.2
Δ Stars	21163	0.02	0.80	-3	4
Δ Fundraising Expenses	31984	-0.81	19.68	-1508.97	1142.413
Δ End of Year Assets	32012	10.57	248.92	-16267.63	12788.54
Δ Program Revenue	31988	0.57	46.83	-4598.29	2572.35
Δ Total Expenses	32010	5.61	229.41	-24086.9	21990.06
Δ Fundraising/Expenses	31983	-0.01	0.06	-.995	.955

Summary statistics are at the charity-year level. Monetary values all normalized to \$2010 and in units of \$100,000.

[4]US BLS data available at https://www.bls.gov/cew/classifications/areas/county-msa-csa-crosswalk.htm. Specifically, the December 2003 CBSA delineations are used since the NCCS data comes from 2000 to 2010.

End-of-year assets and program revenues are sometimes reported as negative. The NCCS KnowledgeBase page gives the following explanations for why this might be the case:

> If there is a bank overdraft in a checking account and no other assets to compensate, then an organization can either report their financial position as showing negative assets or enter the overdraft as an account payable, which is a liability... Some organizations do report negative program service revenues. There are a number of possible reasons. For example, the Lancaster Health Alliance (1998-1999 return) recorded "equity in net earnings of subsidiaries or affiliates" as a category of program service revenue in Part VII of the Form 990. For that year, its losses totalled more than 4.5 million. [5]

Following Okten and Wisebrod (2000), I linearly transform financial variables so they each have a minimum value of at least $1 before log-transforming them for use in regressions.

Charity Navigator does not change ratings for every organization every year, nor does it update ratings on the same day from year-to-year. An organization's rating in a fiscal year where no new rating was posted can be safely assumed to be identical to its most recent previous rating, as that rating will persist on the website until updated. In cases where an organization is rated multiple times in the same fiscal year, I retain only the more recent rating.[5] In the sample, only about 42 percent of organizations have fiscal reporting periods that align with the calendar year. Thus, ratings data is matched to each charity's reporting period by month and then coded as belonging to the year in which the organization's reporting period ends (e.g., Bangor Humane Society received a 3-star rating

[5]I do not have a way of telling which rating may have been responded to, but a large amount of giving occurs toward the end of any given year, with 30% of all annual giving occurring in December [9], and one might intuitively expect donors affected by ratings to use the most up-to-date ratings when giving. As will be discussed in Section 4.1. and pictured in Table 4, over half of these most recent ratings are available to donors for at least four months before the end of the rated organization's fiscal year.

on 11/1/2007, which falls within its reporting period from 5/1/2007 to 4/1/2008, so its FYE 2008 rating is coded as 3 stars).

A charity's score is calculated by summing scores computed by looking at seven equally weighted categories that can be calculated from the charity's Form 990 financial information. Descriptions of these categories and their relationship to score are summarized in Table 3.2.

Table 3.2: Charity Navigator Performance Categories

Performance Category	What it Measures	Effect on Score
Fundraising Efficiency	Amount spent to raise $1	Negative
Fundraising Expenses	% total expenses on Fundraising	Negative
Program Expenses	% total expenses on Programs/Services	Positive
Administration Expenses	% total expenses on Management/General	Negative
Primary Revenue Growth	24-month growth of primary revenue.	Positive
Program Expenses Growth	24-month growth of program expenses	Positive
Working Capital Ratio	# years net liquid assets could sustain spending	Positive

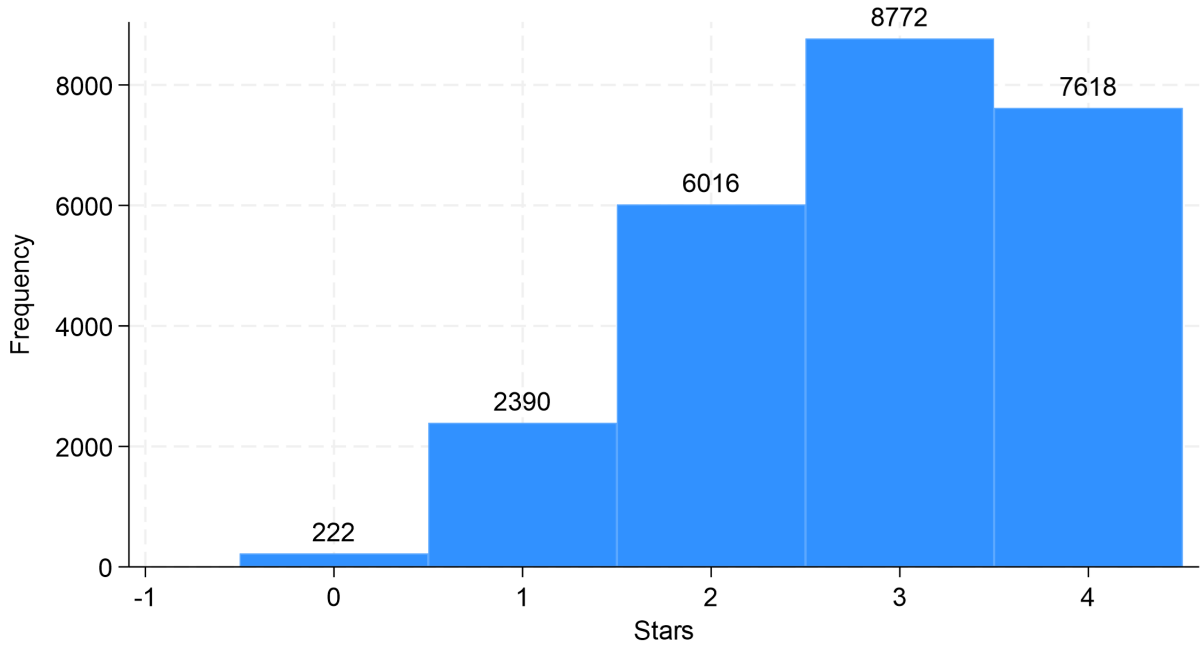
Summary of performance categories listed on Charity Navigator's website in 2002 for their rating system Version 1.0.

Charity Navigator's standards for what level charities should reach in these categories to receive specific scores differ between charity types. For instance, Charity Navigator is harsher to Food Banks with high administrative expenses than Museums, as Food Banks are primarily non-cash operations with little need for overhead spending, and Museums have expensive collections to maintain and expand (the median administrative expenses for these groups also significantly differs from the median among all charities, with Food Banks being much lower and Museums being higher).[6]

[6]More details on Charity Navigator's early rating methodology can be found at

The frequency of star ratings among organizations is shown in Figure 3.1. Charities most commonly receive 3-star or 4-star ratings, and only a small fraction (less than 1%) of ratings are the lowest, 0-star, rating. Figure 3.2 shows the frequency of changes in star ratings. Star ratings are typically stable but will occasionally shift up or down by one with approximately equal likelihood and will rarely make large swings.

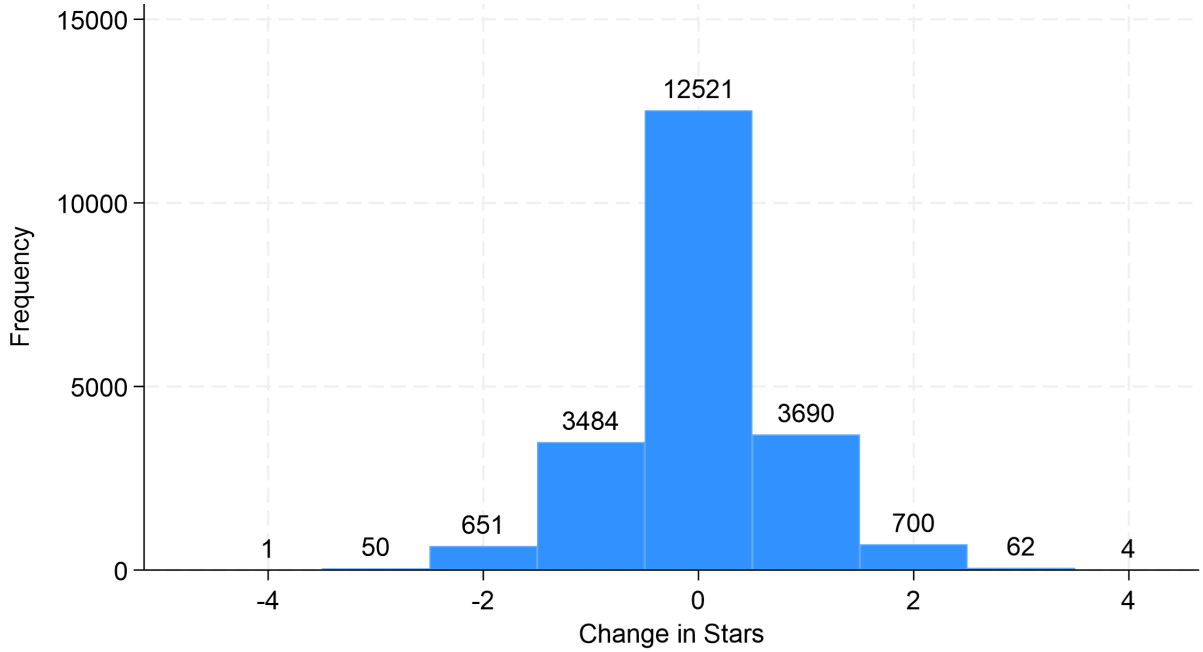
Figure 3.1: Frequency of Star Ratings



Charity-year level frequency of the number of stars organizations received from Charity Navigator. Data comes from a webscrape of historical ratings for 3,882 organizations rated by Charity Navigator between 2002 and 2010.

https://web.archive.org/web/20020806110227/http:/charitynavigator.org/index.cfm/bay/content.view/catid/2/cpid/48.htm.

Figure 3.2: Frequency of Change in Stars



Charity-year level frequency of the year-to-year change in stars organizations received from Charity Navigator. Data comes from a webscrape of historical ratings for 3,882 organizations rated by Charity Navigator between 2002 and 2010.

15
METHODOLOGY

To categorize firms into peer groups, I use the NTEE Core Codes from the NCCS data. These codes were developed by the NCCS to classify nonprofits under one of approximately 400 categories. For example, a professional society called "The Learned Society of Landscape Photographers" would be given an NTEE-CC code of A034, where the major group 'A' stands for Arts, the '03' designates the organization as a professional society, and the final '4' indicates it is a visual arts organization [25]. Following past empirical papers examining the impact of charitable competition, I filter the codes through a series of selection criteria and use the remaining ones to group only nonprofits in the same Metropolitan Statistical Area (MSA). The charities classified must: (1) be local in terms of consumption of output and source of donations, (2) be "reasonably homogeneous" across MSAs, (3) provide services that are substantially distinct from for-profit firms, and (4) receive a nontrivial fraction of their revenues from donations [11, 41]. This is intended to lessen bias from the fact that MSA boundaries can be overly narrow for large organizations with national and international presence and the Internet enables easy transfer of funds across geographical boundaries [11]. Table 4.1 shows the different industry sectors created with associated NTEE codes and the number of separate MSA-based peer groups containing more than one firm associated with each sector.

Table 4.1: Charity Peer Groups

Sector (NTEE Codes)	Number of Peer Groups	Average Firms Per Group
Museums (A50-A57)	33	4.45
Performing arts (A62-A6C)	30	4.53
Community health treatment (E30-E42)	6	2.67
Abuse prevention (I70-I73)	3	2
Employment and vocational training (J20-J33)	5	2.4
Substance abuse prevention and treatment (F20-F22)	1	3
Crime prevention and rehabilitation (I20-I44)	2	4.5
Food pantries and programs (K30-K36)	18	3.28
Homeless Shelters (L40-L41 & P85)	15	3.07
Community Centers (P28)	3	4
Senior centers (P81)	1	2
Total	117	3.83

Charities are considered to share a peer group if both fall into the same sector and are located in the same Metropolitan Statistical Area. Sectors are defined following Thornton (2006). Not all sectors outlined by Thornton (2006) are represented here, as not all had multiple charities rated within the same MSA.

Under this method of grouping peers, 448 organizations have a nonzero number of peers and can form a total of 117 groups with an average of 3.83 firms per group. Summary statistics for this subset of the full sample are displayed in Table 4.2.

Notably, this subsample has lower standard deviations for all variables, though similar means. The largest differences are the subsample's larger means for End of Year Assets, Program Revenue, and the average year-to-year changes in those variables. There are also no negative values reported for any of the financial variables in the subsample.

Table 4.2: Charity Summary Statistics (Organizations with Peers by Sector + MSA)

	N	Mean	SD	Min	Max
Total Contributions	4583	144.01	418.41	0	10495.22
Score	2949	82.29	9.46	48.5	99.83
Stars	2949	2.74	0.95	0	4
Fundraising Expenses	4576	8.86	17.17	0	295.8308
End of Year Assets	4583	716.18	2283.90	.0034794	38429.07
Program Revenue	4577	46.49	98.85	0	1480.352
Total Expenses	4582	190.95	477.89	.0002546	11117.59
Fundraising/Expenses	4576	0.07	0.07	0	.8399328
Δ Total Contributions	3875	3.11	199.25	-5391.689	7390.888
Δ Score	2506	0.28	7.78	-32.41	45.67
Δ Stars	2506	0.03	0.84	-3	4
Δ Fundraising Expenses	3867	-0.92	10.01	-190.2312	241.7054
Δ End of Year Assets	3875	20.70	367.07	-8189.027	12788.54
Δ Program Revenue	3868	0.18	23.92	-574.9048	356.5968
Δ Total Expenses	3874	6.38	124.43	-1084.299	6673.374
Δ Fundraising/Expenses	3867	-0.01	0.05	-.7620522	.5831857

Summary statistics are at the charity-year level. Monetary values all normalized to \$2010 and in units of \$100,000. These are summary statistics for only charities that have at least one peer. Charities are considered to share a peer group if both fall into the same sector and are located in the same Metropolitan Statistical Area. Sectors are defined following Thornton (2006).

A hypothetical ideal randomized controlled experiment that could measure the causal effect of competitor ratings on a charity's contributions might involve acquiring a representative sample of rated charities and grouping them into industries offering substitutable services. Suppose for the sake of simplicity that each peer group was only composed of two firms. Within a random subset of the peer pairs, one firm could be randomly selected to

receive a rating change. Then, charities whose peer firm's rating changed could be compared to charities whose peer firm's rating did not change to uncover the effect of a change in the peer firm's rating.

Even if firms had multiple peers, if only one peer's score changed within each group where peer scores changed, peer groups with a single changer could be compared to groups with no changer. In a more complex scenario where multiple peer scores had to vary, a random subset of charities within randomly selected peer groups could be chosen to receive no rating change while the others in that group receive rating changes, and in other groups, all ratings would remain fixed. Then, charities whose competitor ratings changed more could be compared to those whose competitor ratings changed less or did not change. In this complex case, changes in peer rating becomes a treatment with varying levels of intensity rather than a binary one.

To test whether changes in peer ratings affect the contributions a charity receives, I employ a regression discontinuity design exploiting the fact that a charity's star ratings from Charity Navigator are assigned at specific thresholds of the overall score the organization receives. Since firms likely have (at best) imprecise control over the scores their peers receive, we should not expect firms with peers slightly above or below the cutoff for a particular star rating to be systematically different from one another. The basic model is specified as follows:

$$\begin{aligned} ln(Contributions_{ijt}) = \beta_0 + \beta_1 I(score_{jt} \geq cutoff) + \beta_2 score_{jt} \\ + \beta_3 score_{jt} * I(score_{jt} \geq cutoff) + \eta_{ij} + \mu_t + e_{ijt} \end{aligned} \tag{4.1}$$

where the dependent variable is the natural log of the contributions received by firm i when considered as a pair with peer firm j in fiscal ending year t in 2010 dollars. The main coefficient of interest, β_1, measures the impact of a one-star increase in star rating from peer firm j in fiscal ending year t on i's contributions.[1] η_{ij} denotes firm-peer pair fixed effects, and μ_t denotes fiscal end year fixed effects.

[1]I keep both rating and contributions in the same period in the main specification, because over 50% of newly posted ratings are visible for at least 5 months before the end of an organization's fiscal period, so present-period contributions could plausibly be affected by the rating in the same period. I check the sensitivity of my results to using prior-period competitor ratings and do not find significant differences in the outcome.

RESULTS

Main Results

Table 5.1 gives the RDD estimates of the effect of a change in peer rating on contributions separately for each cutoff and pooled together with all cutoffs centered around zero.[1] Figure 5.1 presents a visualization of the trends in contributions near each star cutoff using the fixed effects and bandwidths from column (3) in Table 5.1.

In the first column where only year and firm fixed effects are included, "treated" observations include firms with peers whose scores are just above the cutoff whether or not those peers were ever also located beneath the cutoff, and "untreated" observations include firms with peers with scores just below the cutoff whether or not those peers ever cross it. Including both year and firm-pair fixed effects (as in column 3-4) eliminates firm-peer pairs where the peer never appears near both ends of the threshold, more precisely isolating the effect of a specific peer crossing on a specific firm rather than the effect of simply having more highly rated peers. Table 5.2 gives an idea of how these regressions are powered by showing how many firms and peer groups had at least one peer observed both above and below each threshold within the bandwidth used for the regressions in column (3) of Table 5.1. Notably, larger bandwidths were needed around both the two-star and one-star cutoffs to include sufficient variation, and too few organizations were at some point observed on both sides of the one-star cutoff to generate a reliable estimate, so results for this cutoff are

[1] This pooled estimate treats the effect as identical around all cutoffs, though it might reasonably be expected that the effect would vary... e.g., a peer firm rising from three to four stars might be more impactful than that peer rising from two to three stars.

not included in the main table.

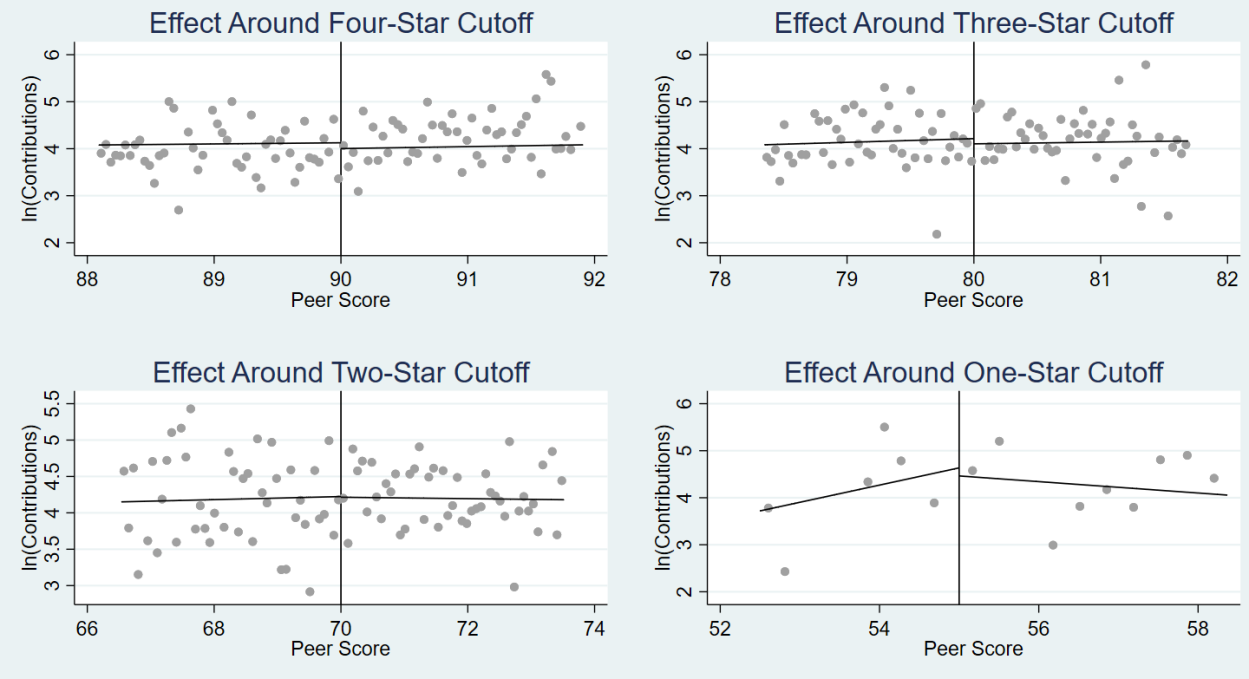
Table 5.1: RDD Estimates of the Effect of a Change in Peer Star Rating on ln(Contributions)

	(1)	(2)	(3)	(4)
4 Stars	.007	-.099***	-.102***	-.099***
	(.016)	(.006)	(.005)	(.005)
3 Stars	-.071*	-.171***	-.131***	-.127***
	(.040)	(.024)	(.010)	(.012)
2 Stars	-.018	.012*	.038**	.046***
	(.022)	(.009)	(.011)	(.008)
Pooled	.004	-.060***	-.065***	-.065***
	(.020)	(.009)	(.009)	(.009)
Year FE	Y	Y	Y	Y
Firm FE	Y	Y	N	N
Peer FE	N	Y	N	N
Firm X Peer FE	N	N	Y	Y
Controls	N	N	N	Y
Observations	21565	21565	21565	20358

Each cell shows an estimate from a separate RD regression where the dependent variable is ln(Contributions) of firm i when considered as a pair with peer firm j in period t and the running variable is the score of peer firm j in period t. Each row shows regressions for a different score cutoff representing a transition to a different star rating assigned by Charity Navigator. The "Pooled" row is created by centering all score cutoffs at 0. Results for 1 Star effect are not shown due to a lack of observations. Robust standard errors in parentheses are clustered at the peer group level. Controls are the prior-period logs of firm i's fundraising expenses, total expenses, end of year assets, and program revenue. All regressions utilize a first-order local polynomial with data driven bandwidth selection from the rdrobust package for Stata.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 5.1: Trends in Contributions Near Cutoffs



Trends in the natural log of contributions to charity i in period t near each score cutoff used by Charity Navigator to determine how many stars a charity is rated. Points shown are sample averages within bins. As in column (3) of Table 5.1, year fixed effects and firm-peer pair fixed effects are controlled for.

When controlling for firm and peer fixed effects, all estimated coefficients of interest in the main model are negative and statistically significant at the .01 level, except for the effect around two stars, which is insignificant when peer fixed effects are not included and positive and significant when they are. The specific magnitude of the coefficients is difficult to interpret. If each firm only had one peer, it would be the local average treatment effect of a firm's partner just barely being in a higher star category rather than the one below it. However, firms may have multiple peers appearing on either side of the cutoff at the same time. The direction of the coefficients is still informative, with a negative coefficient

suggesting that on average firms paired with peers who barely make the cutoff to a three- or four-star rating have lower contributions than when those peers barely miss the cutoff to the higher rating. The positive coefficient on the two-star effect is puzzling, as one would not expect a peer firm rising from one star to two stars to have much of an effect on its peers, much less increasing their contributions.

Table 5.2: Number of Crossings Over Star Cutoffs Within Bandwidths

Cutoff	Bandwidth	Crossing Firms	Firms with Crossing Peer(s)	Peer Groups with Crossing Firm(s)
4 Stars (90)	1.913	48	231	35
3 Stars (80)	1.699	35	212	29
2 Stars (70)	3.527	39	208	33
1 Star (55)	3.472	3	40	3

For each cutoff, this table shows the number of firms observed both above and below that cutoff within the listed bandwidth at least once, the number of firms with at least one peer observed at least once both above and below that cutoff within the listed bandwidth, and the number of peer groups containing at least one firm observed at least once both above and below that cutoff within the listed bandwidth. Cutoffs are the numeric scores used by Charity Navigator to assign its star ratings. 4 Stars are assigned at a score of 90 or above, 3 stars at a score from 80 to below 90, 2 stars at a score from 70 to below 80, 1 star at a score from 55 to below 70, and zero stars at a score below 55. The second column shows the size of the bandwidth used on each side of the score cutoff in the RD regressions in column (3) of Table 5 as determined using data-driven bandwidth selection via the rdrobust package in Stata. For reference, the total number of firms in the sample is 448, and they form 117 peer groups in total.

A negative relationship between competitor ratings and own contributions would be consistent with charitable substitution taking place, but other mechanisms could be at play. In particular, there is evidence that organizations are not randomly sorting around the peer

star cutoffs, which could also explain the puzzling two-star result.[2] Though peers may not be manipulating each other's scores, there may be systematic differences in the characteristics of organizations likely to have a peer on one side of the cutoff in a given period rather than the other. This concern is discussed in detail in the next section of this chapter.

Plausibility/Limitations of RD Design

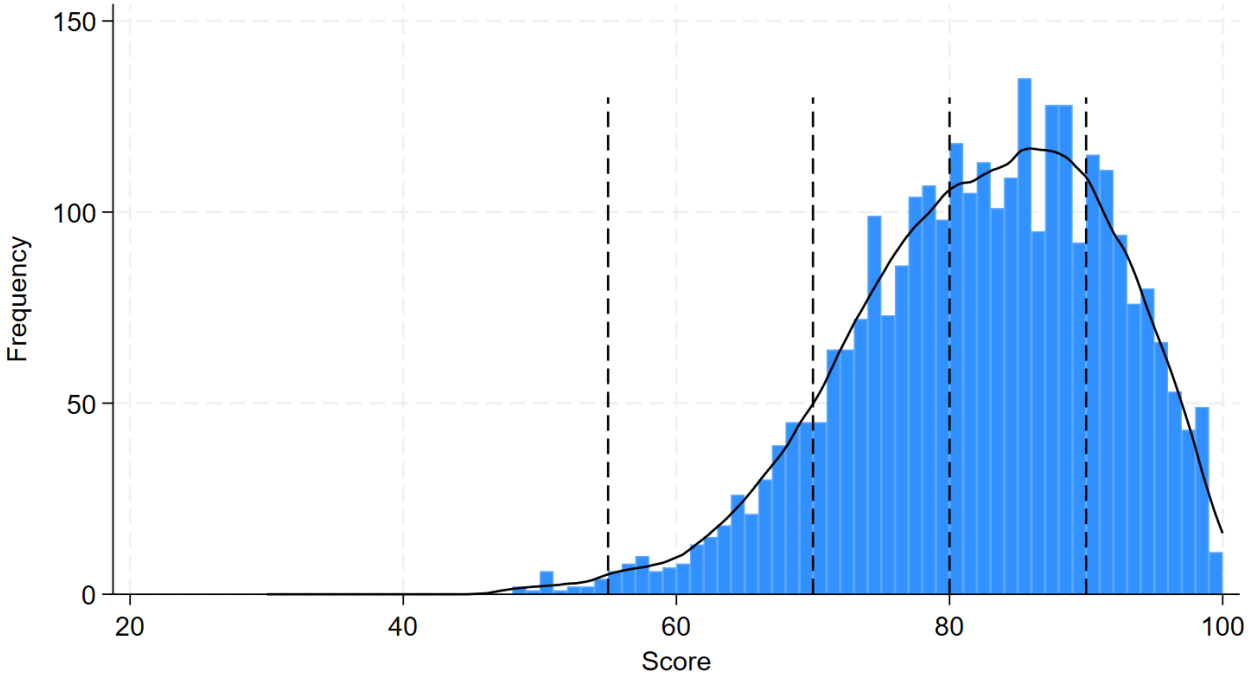
The identifying assumptions of the RD model are that there is random sorting of organizations close to the cutoffs for different star thresholds and that other variables which influence contributions do not experience discontinuous jumps at the thresholds. Intuitively, it is highly implausible that firms have precise control over the ratings of their peers. An organization's score is based on financial data reported by that nonprofit itself, and a peer organization would only indirectly be able to affect that nonprofit's contributions, fundraising, expenses etc. (e.g., perhaps it could try to induce its peer to fundraise more by launching its own fundraising campaign, but this would not translate to precise control over peer fundraising expenditures).

It is difficult to empirically test whether organizations may be manipulating the scores of their peers in peer groups larger than two firms, but I am able to test for the possibility that charities manipulate their own scores around the cutoffs using McCrary density tests

[2]That being said, the positive two-star effect is not incomprehensible at face value: one could imagine situations where peer groups containing a greater proportion of lower-rated firms react to peers rising one star by fundraising more, and if donors don't substitute much towards the higher-rated charities, the overall effect may be an increase in contributions for the charities fundraising more in response to the peer rating change.

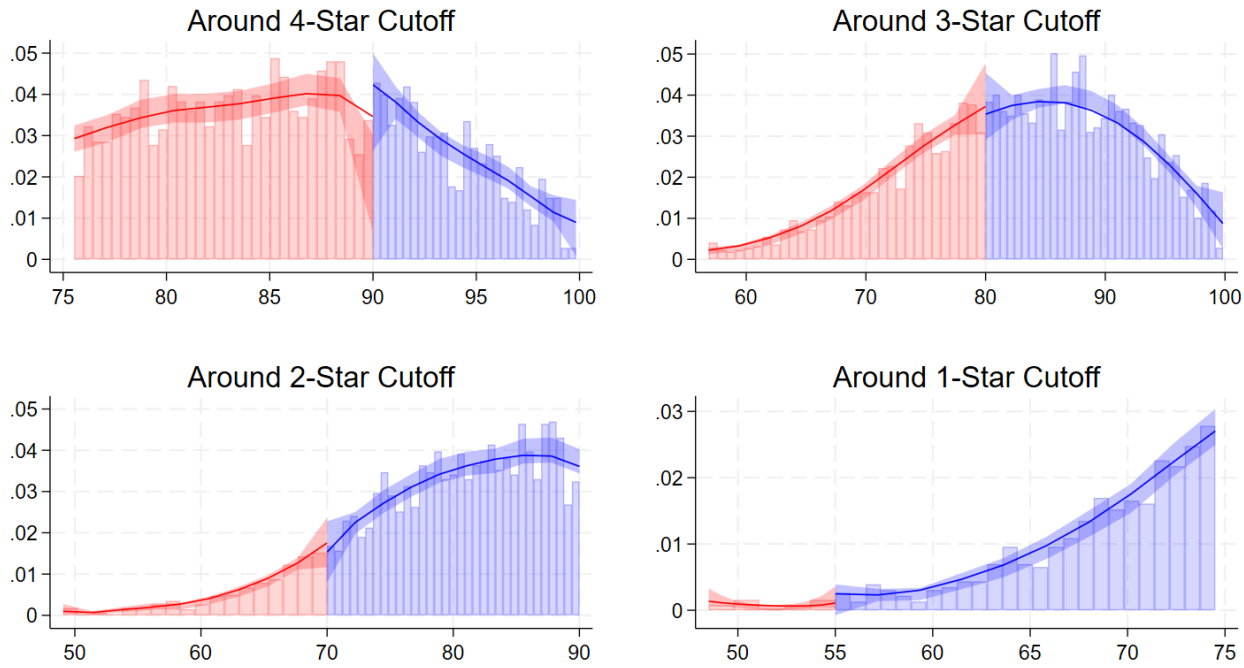
(McCrary, 2008). Figure 5.2 shows the distribution of individual scores and where the star cutoffs are for the subsample used in the regressions, and Figure 5.3 shows the results of manipulation tests around each of those four cutoffs.

Figure 5.2: Frequency of Own Score (Subsample)



Charity-year level frequency of the numerical score organizations received from Charity Navigator. Dashed lines at the 55, 70, 80, and 90 marks indicate the score threshods used to determine how many stars (0-4) an organization is rated (e.g., scores beyond the rightmost line correspond to 4-star charities). Data comes from a webscrape of historical ratings for 448 organizations rated by Charity Navigator between 2002 and 2010. Organizations were selected for inclusion from my full sample of 3,882 organizations if they had at least one peer. Nonprofits are considered to share a peer group if both fall into the same sector (defined following Thornton, 2006) and are located in the same Metropolitan Statistical Area.

Figure 5.3: Manipulation Tests for Own Score (Subsample)



Visualizations for formal tests of the ability of charities to manipulate the scores around the thresholds used for determining the stars they receive from Charity Navigator based on McCrary (2008). Data comes from a webscrape of historical ratings for 448 organizations rated by Charity Navigator between 2002 and 2010. Organizations were selected for inclusion from my full sample of 3,882 organizations if they had at least one peer. Nonprofits are considered to share a peer group if both fall into the same sector (defined following Thornton, 2006) and are located in the same Metropolitan Statistical Area.

The null hypothesis that there was no manipulation of organization's own scores can only be rejected around the 4-star threshold at the 0.05 level. Even so, the fact that it is difficult to find evidence that organizations can nonrandomly sort themselves into star categories near the majority of the cutoffs lends credibility to the even more plausible assumption that charities cannot precisely manipulate the scores of their peers close to the cutoffs. Additionally, if manipulation of one's own score is possible, one would expect

firms to be able to exploit this ability to manipulate their own scores to counteract possible manipulation of their scores by peers.

I next check for the continuity of various variables shown to be related to a nonprofit's contributions across the star rating thresholds. Prior literature has identified prior-period fundraising [31, 46], expense ratios [31, 42–44, 46], and total assets [42, 43] as determinants of contributions. Some literature has failed to find evidence that program service revenue crowds out donations [31], but recent charitable ratings literature has continued to include it to account for the possibility of such an effect (Yörük, 2016).[3] Results for these continuity checks are shown in Table 5.3.

The results from these continuity checks raise severe concerns for the ability to interpret the results of the main model as identifying a causal relationship between peer ratings and contributions. Present-period changes in peer score cannot plausibly impact past values of charity financial variables, so statistically significant discontinuities in past period financial variables around the peer score cutoffs imply nonrandom sorting of organizations around the cutoffs. These discontinuities are also in directions that could explain the observed results. For instance, at the four-star cutoff, a large negative relationship in an organization's prior-period fundraising is present. Holding other factors constant, lower levels of prior-period fundraising are associated with lower levels of present-period contributions. The puzzling

[3]Other key determinants of charitable contributions include shifts in the incomes of recipients of charitable goods/services (Becker, 1974), industry, age, changes in tax laws [31], and donor characteristics like personal income, net capital gains, religious affiliation, volunteerism, race [19], and subjective dispositions [29]. I do not have data on many of these variables, but they should be completely unrelated to Charity Navigator's ratings.

two-star result from Table 5.1 might similarly be explained by these continuity tests, as increases in prior-period fundraising and end of year assets are positively associated with increases in present-period contributions.

Table 5.3: RD Continuity Checks for Prior-Period Financial Variables

	$\ln(\text{Fundraising})_{i,t-1}$	$\ln(\text{Expenses})_{i,t-1}$	$\ln(\text{Assets})_{i,t-1}$	$\ln(\text{ProgramRevenue})_{i,t-1}$
4 Stars	-.962***	-.013***	-.002***	-.047***
	(.062)	(.003)	(.004)	(.005)
3 Stars	-1.058***	.028***	.051***	.003
	(.050)	(.006)	(.004)	(.006)
2 Stars	.134**	-.006	.025***	.026**
	(.066)	(.004)	(.007)	(.007)
Year FE	Y	Y	Y	Y
Firm X Peer FE	Y	Y	Y	Y
Observations	21443	21499	21499	21443

Each cell shows an estimate from a separate RD regression with peer score of firm j in period t as the running variable. Each column shows the estimates at different star thresholds on a separate dependent variable: a firm's own prior-period logged fundraising expenses, total expenses, end of year assets, and program revenue. Robust standard errors in parentheses are clustered at the peer group level. All regressions utilize a first-order local polynomial with the same bandwidths listed in Table 5.2 used in the regressions in column (3) of Table 5.1.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Additional Robustness Checks

The sensitivity of the main results to the use of different local polynomials and bandwidths can be seen in Table 5.4 and Table 5.5.

Table 5.4: Polynomial Sensitivity of RDD

	Linear	Quadratic	Cubic
4 Stars	-.102***	-.071***	-.097***
	(.005)	(.014)	(.016)
3 Stars	-.131***	-.145**	-.143***
	(.010)	(.017)	(.021)
2 Stars	.038***	-.035***	-.004
	(.011)	(.011)	(.017)
Pooled	-.065***	-.091***	-.071***
	(.009)	(.012)	(.018)
Year FE	Y	Y	Y
Firm X Peer FE	Y	Y	Y
Observations	21565	21565	21565

Each cell shows an estimate from a separate RD regression where the dependent variable is ln(Contributions) of firm i when considered as a pair with peer firm j in period t and the running variable is the score of peer firm j in period t. Each row shows regressions for a different score cutoff representing a transition to a different star rating assigned by Charity Navigator. The "Pooled" row is created by centering all score cutoffs at 0. Robust standard errors in parentheses are clustered at the peer group level. Regressions in each column utilize a different local polynomial. All regressions use the same bandwidths listed in Table 6 used in the regressions in column (3) of Table 5.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.5: Bandwidth Sensitivity of RDD

	Original BW	Half BW	Double BW
4 Stars	-.102***	.010	-.086***
	(.005)	(.005)	(.013)
3 Stars	-.131***	-.343***	-.061***
	(.010)	(.012)	(.010)
2 Stars	.038**	.006	-.008
	(.011)	(.006)	(.012)
Pooled	-.065***	-.064***	-.045***
	(.009)	(.009)	(.009)
Controls	N	N	N
Year FE	Y	Y	Y
Firm X Peer FE	Y	Y	Y
Observations	21565	21565	21565

Each cell shows an estimate from a separate RD regression where the dependent variable is ln(Contributions) of firm i when considered as a pair with peer firm j in period t and the running variable is the score of peer firm j in period t. Robust standard errors in parentheses clustered at the peer group level. All regressions utilize a first-order local polynomial. Each column shows regressions for different bandwidths: the original data driven bandwidths computed with the rdrobust package for Stata (which are unique to each star threshold and listed in Table 6), half of this bandwidth, and double this bandwidth.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The three-star result appears the most stable despite changes in polynomial and bandwidths, while the two-star result is quite sensitive, switching signs under a quadratic polynomial and losing significance when using a cubic or changing the bandwidth. The four-star estimate is fairly stable under different polynomials but becomes positive and statistically insignificant at half the original bandwidth.

Though the average peer group size is about 4 firms, groups in the sample can be as large as 39 firms. Because the same firm might appear in the regression multiple times in the same period if it had more than one peer near the threshold in that period, it might be the case that the results are primarily driven by the inclusion of large peer groups. Table 10 shows the sensitivity of the main results to changes in the sample used for estimation based on peer group size.

In general, the two-star result is not robust to excluding the largest peer groups, while the three- and four-star estimates retain their direction and significance for at least the 0.1 level throughout. The significance of the three-star rating effect is lower when only including groups with single-digit sizes compared to only including groups of double-digit sizes. One possible explanation for this is heterogeneity in the treatment effect. Peer ratings might be more important in larger peer groups if it is more difficult to independently evaluate and compare organizations with more peers. However, it might be the case that there is more variation in larger peer groups that gives the model more power.

Table 5.6: Robustness of RD Estimates to Modifying Sample Based on Peer Group Size

	Drop Largest Group	Drop 3 Largest Groups	Group Size < 10	Group Size ≥ 10
4 Stars	-.148***	-.181***	-.116***	-.065***
	(.010)	(.017)	(.019)	(.004)
3 Stars	-.172***	-.270***	-.075*	-.140**
	(.020)	(.026)	(.026)	(.010)
2 Stars	.034**	.081	.066	-.016
	(.015)	(.024)	(.020)	(.011)
Year FE	Y	Y	Y	Y
Firm X Peer FE	Y	Y	Y	Y
Observations	13048	7610	6299	15266

Each cell shows an estimate from a separate RD regression where the dependent variable is ln(Contributions) of firm i when considered as a pair with peer firm j in period t and the running variable is the score of peer firm j in period t. Different columns include/exclude organizations depending on the size of their peer group. Robust standard errors in parentheses clustered at the peer group level. All regressions utilize a first-order local polynomial with the same bandwidths listed in Table 6 used in the regressions in column (3) of Table 5.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To test whether the main results and the observed discontinuities in prior-period financial variable values are influenced by organizing firms in peer groups of varied sizes, I use the same model with a method for grouping firms into pairs so each peer group is of size two. Firms are matched with their "closest peer," defined as the firm in the same peer group (by Sector + MSA) with the most similar value for average end of year assets over the sample period.[4] Results for these regressions are reported in Table 5.7 and Table 5.8.

[4]Selecting "closest peers" using average total expenses yields similar results.

Table 5.7: RDD Estimates of the Effect of a Change in Closest Peer Star Rating on ln(Contributions)

	(1)	(2)
4 Stars	-.243**	-.056
	(.037)	(.029)
3 Stars	.037	.030
	(.036)	(.035)
2 Stars	-.590***	-.696***
	(.032)	(.022)
Controls	N	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	2695	2521

Each cell shows an estimate from a separate RD regression with score of firm i's closest peer in period t as the running variable. Firm i's closest peer is defined as the firm in the same peer group with the closest average value of End of Year Assets over the sample period. Controls are the prior-period logs of firm i's fundraising expenses, total expenses, end of year assets, and program revenue. All regressions utilize a first-order local polynomial with data driven bandwidth selection from the rdrobust package for Stata.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.8: Closest Peer RD Estimates for Prior-Period Financial Variables

	$\ln(\text{Fundraising})_{i,t-1}$	$\ln(\text{Expenses})_{i,t-1}$	$\ln(\text{Assets})_{i,t-1}$	$\ln(\text{ProgramRevenue})_{i,t-1}$
4 Stars	-2.096***	.067	-.000	-.109***
	(.194)	(.013)	(.021)	(.022)
3 Stars	-.933***	.060***	.027	-.057
	(.146)	(.014)	(.022)	(.018)
2 Stars	.002	.104***	.020	-.448***
	(.078)	(.018)	(.027)	(.031)
Year FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Observations	2685	2685	2685	2685

Each cell shows an estimate from a separate RD regression with score of firm i's closest peer in period t as the running variable. Firm i's closest peer is defined as the firm in the same peer group with the closest average value of End of Year Assets over the sample period. Each column shows the estimates at different star thresholds on a separate dependent variable: a firm's own prior-period logged fundraising expenses, total expenses, end of year assets, and program revenue. Robust standard errors in parentheses are clustered at the firm level. All regressions utilize a first-order local polynomial with the same bandwidths as in column (1) of Table 11.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As in the main results and continuity checks, discontinuities in the values of prior-period financial variables which could explain the changes in present-period contributions are present around the cutoffs. The results of the primary regressions and continuity checks are not merely a byproduct of how pairing firms with multiple peers leads the estimates to be weighted. This provides further evidence of nonrandom sorting around the cutoffs, suggesting that firms systematically differ when their peers change in star rating, even near the thresholds for changing rating, in ways that are not caused by the change in star rating.

CONCLUSION

Donations to nonprofits often take place in conditions where there is uncertainty as to the quality and value added of charitable output. Nonprofit rating organizations like Charity Navigator may play an informational/disciplinary role in the market for charitable contributions through publicizing ratings on different organizations. When those organizations offer substitutable services, these ratings might be used to compare competing charities and inform the decision to donate. This paper finds evidence that if a charity's peers rise to the three- or four-star level, that charity's contributions fall in that period. Though this is consistent with Charity Navigator ratings being used to make donation decisions between charities providing substitutable services, it may be more likely that the results are driven nonrandom sorting around the cutoffs, as present-period peer star rating is also related to past-period values in one's own financial variables with sizes/directions that could explain the main results.

This investigation suffers from several further limitations. First, ratings data for charities rated between 2002 and 2010 could not be scraped from Charity Navigator's website if those organizations no longer exist. If there is some relationship on average between an organization's sensitivity to peer ratings changes and its likelihood of existing for a longer period, the estimated effect of a peer rating change could be biased in either direction depending on whether sensitive organizations are more/less likely to persist. There may be other ways of acquiring this missing data that would allow this limitation to be surpassed in future work. The financial data from the National Center for Charitable Statistics is also

missing for some charities in some years for unknown reasons, so the validity of these results must be caveated as applying specifically to the non-missing data. The generalizability of these findings may also be limited because (1) charities rated by Charity Navigator are fairly large and may not be representative of all charities for which methods of information sharing and peer comparison may be available, and (2) the choice of peer group construction in this paper excludes nonprofit organizations competing in larger regional, national, or international markets for donations.

Finally, it should be noted that evidence of contributions responding to changes in peer rating, even if causal, only supports the idea that ratings organizations like Charity Navigator help allocate donated dollars to more effective uses insofar as the ratings are effective proxies for organizational quality. If donors are responsive to peer ratings, but peer ratings are overly simplistic or deceptive measures of relative quality, rating agencies could be ineffective or harmful to the allocation of donations in society.[1] However, the creation and refining of mechanisms by which substitutable charities can be identified and compared may increase the ability of charitable markets to allocate donations to the cost-effective provision of public goods. Thus, innovations in the space of charity ratings will likely require ongoing examination of the effects of ratings on donors and their merit as assessments of charity quality.

[1]One nonprofit law expert has cited shortcomings in charity rating systems as sufficient cause to advocate that "the making and dissemination of ratings [from watchdog agencies] ought to be abolished" [23]. In a softer critique, Stinn (2019) laments what he views as a counterproductive "obsession" with overhead ratios from both donors and watchdogs, as pressure to keep overhead low can starve nonprofits of resources with which to implement their programs.

REFERENCES CITED

[1] About us. Charity Navigator. Retrieved August 15, 2023, from https://www.charitynavigator.org/about-us/.

[2] BBB standards for charity accountability. Give.org. Retrieved August 15, 2023, from https://give.org/donor-landing-page/bbb-standards-for-charity-accountability.

[3] Internet archive: About IA. Internet Archive. Retrieved December 22, 2023, from https://archive.org/about/.

[4] (2005). Rating the raters: An assessment of organizations and publications that rate/rank charitable nonprofit organizations.

[5] (2006). Negative numbers for financial variables (technical note).

[6] (2020). Charity watchdogs: Ignore them at your own risk. Technical report, Grey Matter Research and Harmon Research.

[7] (2021). Health of the u.s. nonprofit sector.

[8] (2022). Comparing charities: How big is the difference?

[9] (2023). The ultimate list of charitable giving statistics for 2023.

[10] Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401):464–477. Publisher: [Royal Economic Society, Wiley].

[11] Bose, B. (2015). Effects of nonprofit competition on charitable donations. *University of Washington*.

[12] Chen, G. (2009). Does meeting standards affect charitable giving? an empirical study of new york metropolitan area charities. *Nonprofit Management and Leadership*, 19(3):349–365. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nml.223.

[13] Chhaochharia, V. and Ghosh, S. (2008). Do charity ratings matter? Unpublished Manuscript, Florida Atlantic University.

[14] Cnaan, R. A., Jones, K., Dickin, A., and Salomon, M. (2011). Nonprofit watchdogs: Do they serve the average donor? *Nonprofit Management and Leadership*, 21(4):381–397. Publisher: John Wiley & Sons, Ltd.

[15] Dennis, C. (2021). Charity ratings compared: Charity navigator vs. CharityWatch vs. guidestar vs. GreatNonprofits. Section: Insights.

[16] Deryugina, T. and Marx, B. M. (2021). Is the supply of charitable donations fixed? evidence from deadly tornadoes. *American Economic Review: Insights*, 3(3):383–398.

[17] Ebarb, T. S. (2019). Nonprofits fail – here's seven reasons why.

[18] Filiz-Ozbay, E. and Uler, N. (2019). Demand for giving to multiple charities: An experimental study. *Journal of the European Economic Association*, 17(3):725–753.

[19] Gittell, R. and Tebaldi, E. (2006). Charitable giving: Factors influencing giving in u.s. states. *Nonprofit and Voluntary Sector Quarterly*, 35(4):721–736. Publisher: SAGE Publications Inc.

[20] Gordon, T. P., Knock, C. L., and Neely, D. G. (2009). The role of rating agencies in the market for charitable contributions: An empirical test. *Journal of Accounting and Public Policy*, 28(6):469–484. Publisher: Elsevier.

[21] Grant, L. (2017). Strategic responses to third-party ratings: Theory and evidence from charities.

[22] Grant, L. E. and Potoski, M. (2015). Collective reputations affect donations to nonprofits. *Journal of Policy Analysis and Management*, 34(4):835–852. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.21868.

[23] Hopkins, B. R. and Kirkpatrick, A. M. (2013). *The Law of Fundraising*. Wiley, 5th edition edition.

[24] Johnson, L. (2021). Top charity watchdogs nonprofits should know.

[25] Jones, D. (2019). National taxonomy of exempt entities (NTEE) codes.

[26] Krasteva, S. and Yildirim, H. (2016). Information, competition, and the quality of charities. *Journal of Public Economics*, 144:64–77.

[27] Ling, Q. and Neely, D. G. (2013). Charitable ratings and financial reporting quality: Evidence from the human service sector. *Journal of Public Budgeting, Accounting & Financial Management*, 25(1):69–90. Publisher: Emerald Publishing Limited.

[28] McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.

[29] Neumayr, M. and Handy, F. (2019). Charitable giving: What influences donors' choice among different causes? *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 30(4):783–799. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 4 Publisher: Springer US.

[30] Null, C. (2011). Warm glow, information, and inefficient charitable giving. *Journal of Public Economics*, 95(5):455–465.

[31] Okten, C. and Weisbrod, B. A. (2000). Determinants of donations in private nonprofit markets. *Journal of Public Economics*, 75(2):255–272.

[32] Prychitko, D. L. and Boettke, P. J. (2004). Is an independent nonprofit sector prone to failure? an austrian school analysis of the salamon paradigm and the lohmann challenge. SSRN Scholarly Paper.

[33] Reinstein, D. (2010). Substitution among charitable contributions: An experimental study. *Unpublished Manuscript*.

[34] Reinstein, D. A. (2011). Does one charitable contribution come at the expense of another? *The B.E. Journal of Economic Analysis & Policy*, 11(1). Publisher: De Gruyter.

[35] Sloan, M. F. (2009). The effects of nonprofit accountability ratings on donor behavior. *Nonprofit and Voluntary Sector Quarterly*, 38(2):220–236. Publisher: SAGE PublicationsSage CA: Los Angeles, CA.

[36] Stinn, J. (2019). Donors need to stop pressuring nonprofits to pinch pennies.

[37] Stork, D. and Woodilla, J. (2008). Nonprofit organizations: An introduction to charity rating sources and cautions in their use. *The International Journal of Applied Management and Technology*, 6(4):5–28.

[38] Szper, R. (2013). Playing to the test: Organizational responses to third party ratings. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 24(4):935–952. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 4 Publisher: Springer US.

[39] Szper, R. and Prakash, A. (2011). Charity watchdogs and the limits of information-based regulation. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 22(1):112–141. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 1 Publisher: Springer US.

[40] Taylor, H. (2021). Wise giving wednesday: Twentieth anniversary of BBB wise giving alliance.

[41] Thornton, J. (2006). Nonprofit fund-raising in competitive donor markets. *Nonprofit and Voluntary Sector Quarterly*, 35(2):204–224. Publisher: Sage PublicationsSage CA: Thousand Oaks, CA.

[42] Tinkelman, D. (1998). Differences in sensitivity of financial statement users to joint cost allocations: The case of nonprofit organizations. *Journal of Accounting, Auditing & Finance*, 13(4):377–393. Publisher: SAGE Publications Inc.

[43] Tinkelman, D. (1999). Factors affecting the relation between donations to non-for-profit organizations and an efficiency ratio. *Research in Governmental and Nonprofit Accounting*, 10:135–161.

[44] Tinkelman, D. and Mankaney, K. (2007). When is administrative efficiency associated with charitable donations? *Nonprofit and Voluntary Sector Quarterly*, 36(1):41–64. Publisher: SAGE Publications Inc.

[45] Weinstein, L. (2024). Research guides: The nonprofit sector in the united states: A resource guide: Form 990.

[46] Weisbrod, B. A. and Dominguez, N. D. (1986). Demand for collective goods in private nonprofit markets: Can fundraising expenditures help overcome free-rider behavior? *Journal of Public Economics*, 30(1):83–96.

[47] Yörük, B. K. (2016). Charity ratings. *Journal of Economics & Management Strategy*, 25(1):195–219. Publisher: John Wiley & Sons, Ltd.