

Really, Don't Tell Me: Voluntary Contributions and the Effects of Watchdog Ratings

*Laura Ellyn Grant*¹

*Work in progress: Do not cite — Comments welcome
April 21, 2010*

Abstract

Charities provide public benefits that are difficult to measure and often unverifiable for the donors, creating a one-sided information deficit. To diminish this information asymmetry, watchdog organizations provide donors with ratings of charities based on each charity's financial performance. The information can increase donor awareness of individual charity performance and provide a metric of relative performance.

Watchdog ratings, ranging from 0- to 4-stars in this study, likely affect donor choice, changing and redistributing donations among charities. This research measures this informational effect on contributions for over 5,400 charities in the US. The data contain eight years of ratings from the largest third-party charity evaluator and eight additional years of tax data prior to the ratings. The first approach uses differences within-charity after the information is available. The second and pivotal tactic calculates, using identical methodology of the evaluator, unpublished scores of each charity from financial data before being rated. This approach isolates the effect of observing the rating by creating a counterfactual of unpublished information to compare with the published. The third specification adds regression discontinuity: star-rating breakpoints have an underlying continuous distribution of scores. Charities with similar scores but a one-star difference in ratings have distinct outcomes in contributions.

Counter to expectation, observed ratings *reduce* contributions, regardless of star rating, by 3 to 30%, where annual donations average \$18 million per charity per year. The model and discussion describe how ratings may lead to donor scrutiny and also how financial information relates to goods and service output. These ideas demonstrate potential unintended consequences of shifting donations based on efficiency rather than effectiveness. The effects are particularly germane to the relatively-smaller charities and those with far-removed goods, such as the environment or foreign aid.

Acknowledgements: Matthew Kotchen, Charlie Kolstad, Paulina Oliva, James Andreoni. Doug Steigerwald and the UCSB Econometrics Research Group. UCSB Environmental Economics Lunch Group. Funding from NSF grant 0114437 and UC Regents Fellowship. This is a draft; feedback and notifying of possible errata are most welcome.

JEL: H41 - Public Goods, D82 - Asymmetric and Private Information

¹ lgrant@bren.ucsb.edu
Bren School of Environmental Science & Management
University of California
4420 Donald Bren Hall, Santa Barbara, CA 93106-5131

1 Introduction

Charitable giving is a large sector in the US economy, with private donations of over \$300 billion in 2008 (~2.1% of US GDP), despite numerous market failures. A market functions well under particular conditions; one aspect is accurate and complete information about the product. Charities provide a diverse array of goods and services, ranging from educational outreach to urban parks, low-income healthcare to land conservation. Yet a commonality among these organizations is diffuse products that have less tangible benefits than typical market goods. Charitable objectives are perceived by donors as inherently positive, yet using celebrity galas for fundraising is criticized as wasteful and, at worst, reports of donation embezzlement undermine charity status.² These problems stem from lack of information available to donors – the inability to discern and validate characteristics of charities – and, as theoretical literature posits, impede efficient market functions.³ When donors are unable to distinguish charitable products and services, the market cannot appropriately allocate donations to deserving charities.

In order to provide oversight and indicate charity credibility, independent watchdog groups have designed financial ratings for the nonprofit charity market. The evaluators calculate this informational rating using various ratios of operating costs and the donations received. The ratings can serve two major purposes: alleviating some information asymmetry for each rated charity and creating an index for comparison between charities.⁴ Whether or not the ratings succeed is the focus of this study.

² This research abstracts from the problems of free-ridership typical of public good provision, focusing instead on information asymmetry. However, donor-side demand and characteristics are explored below, including reference to private provision of public goods from altruism, warm glow, social pressure, etc.

³ Thorough knowledge of charitable activities is unfeasible. There are nearly 1 million registered public charities and 175 additional tax-exempt organizations recognized by the IRS every day. This is approximately one organization for every 169 Americans.

⁴ Financial efficiency is not the only influence on donations; other motivators are accountability and charity impact. However, social impact is exceedingly difficult to measure and not

Herein, I model the demand for charity under asymmetric information and also the changes with the introduction of ratings: the information effect. The framework gives intuition and motivates the empirical design. Then using data from the nation's largest evaluator of charities, this research measures the reduced-form outcome of ratings by analyzing the effects on charitable donations from the levels of ratings, changes in ratings, and ratings in years before and after publication.

This information likely induces shifts in contributions, rewarding organizations with good outcomes and punishing those with poor ratings. Publishing the ratings to the general public could increase overall donations through appreciation of rating availability or could lower aggregate contributions as a result of scrutiny. The charities may also respond to the information. They can self-publicize good ratings to enhance reputation and garner contributions. In response to the feedback of negative ratings, organizations may change behavior to improve future ratings. Charities can respond through either performance mechanisms – changing operations toward financial efficiency and reorienting missions toward donor demands – or strategic reporting mechanisms – changing the accounting on tax forms.⁵ The charity-side effects are enveloped into the demand response.

The ratings indicate efficient organizations to donors, but surprisingly, the effect of this information is currently unidentified.⁶ With hundreds-of-billions-of-dollars in potential donations at stake annually in the United States, determining the

commensurate between charities. We assume that financial efficiency does indicate something about charity reputation and quality. These issues are discussed further in the conclusion.

⁵ A follow-up paper will also test this concept, if the charities are using the feedback and strategically improving their financial status, and thus the subsequent ratings, over time. Perverse incentives for 'gaming the system' – will need to provide evidence. Some charities are audited each year, but I am not sure these are available. Also, there may be a pre-emptive effect if charities know that operational expenses are a factor for donors. These issues will be dealt with in the next paper.

⁶ Impact of charity ratings are not known, with the exception of one unpublished research paper, which will be reviewed, and a handful of general-audience articles. There are many studies that address other aspects of charities, including government grant crowding-out/-in and competition.

marginal effect of this information is crucial. This research sets up a basic model of demand for risky contributions and donor response to improved information through ratings in the charity market. The empirical work measures the information effect of ratings on contributions to charities. The research uses complete data from the largest third-party rating organization, ratings of more than 5,400 large charity organizations longitudinally through eight years. I measure the value of information by exploiting the within-charity variations in ratings, trend of ratings, and variation in timing of being rated.

The full details are given below; however, the treatment design is the pivotal characteristic of the dataset. Ratings are deterministic calculations from publicly available tax data. Using historical IRS financial data for eight additional years, I calculate unobserved charity ratings previous to publication using identical methodology of the evaluator. Organizations are added regularly through time and then rated each subsequent year, to create a growing panel for analysis. This feature allows identification of time trends affecting donor behavior beyond the ratings and provides for tests of selection into the rated group. Then, annual contributions are compared before and after the information becomes available, isolating the effect of the rating publication. The calculated unobserved ratings act as a counterfactual treatment, controlling for idiosyncrasies of the rating.

Despite information's typical role of increasing demand, observed ratings reduce contributions, regardless of star rating, by 3 to 30%, where annual donations average \$18 million per charity. The conclusion describes an underlying cause for the declines, that ratings may induce donor scrutiny. Discussion also covers how financial information relates to the true mission of charities: goods and service output. These ideas demonstrate potential unintended consequences of shifting donations based on efficiency rather than effectiveness. The effects are also discussed in reference to size and category of charities. Additionally, this research gives a sense of the role of information policies as an intervention within the charity market, as a whole.

The next section has background and literature. Donor demand for charity is modeled in section three. Data descriptions are in section four, followed by a section on methods and results. The final section concludes with discussion.

2 Who is the informed donor?

Motives for giving are diverse and complex. Collecting each reason would be prohibitive, as nearly 60 percent of US households give annually and another 25 percent occasionally (Indiana U survey).⁷ Certainly income, educational levels, personal interests, and political beliefs have roles, but donors come from all these spectrums. Giving USA, a nonprofit research group through University of Indiana, performs an annual survey of donor types and contribution behavior – who donates and how much. Altruism seems primary: general stated motivations are ‘helping society’ or ‘giving to a worthy cause’.⁸ Moral obligation and reputation (Tullock 1966), sense of paying back, and guilt (DellaVigna *et al* 2009) are also cited. These reasons interplay in an impure public goods sense, combining private benefits with the value of charitable goods provided. Vesterlund (2006) reviews donor motivation with a tour of the related economic literature. She reconsiders the seminal work of government crowd-out of public goods, grounded in the public versus private benefits of giving (Bergstrom *et al* 1986, Andreoni 1989 & 1990). She also provides evidence from experiments wherein donor types and motivations are tested using carefully constructed contexts of donation opportunities.

Except in the laboratory setting, donating also involves a cost to the donor; the act is giving away money. Thus, another factor is tax deductibility – the subsidies on donations financed from lost tax revenue. If contributions were a perfect substitute for paying taxes, philanthropy would simply be voting for public

⁷ http://www.shanghai.gatech.edu/academics/ECEMS/CoulterArticles/11_householdsdonate.pdf accessed 3/15/10 or <http://www.philanthropy.iupui.edu/Research/copps/panelstudy.aspx>

⁸ www.forbes.com/2008/11/19/americas-largest-charities-pf-charities08-cx_wb_1119charity_land.html accessed March 11, 2010.

services and goods with a dollar rather than through state representatives. However, only a portion of donations is subsidized by the government, at the rate of each household income tax bracket. Ranging over incomes, the donor's loss is between 65 and 100%. Literature shows donor awareness of deductibility (Prince & File 1994) and extensively confirms response to changes in charity tax laws, although with conflicting outcomes (Randolph 1995 and Auten *et al* 2002).

Added costs come from the search for sound charities, often termed due diligence. However because donations are a small portion of household budget each year,⁹ individuals likely put minimal effort into searching for an acceptable charity. Beyond their college *alma mater* or local religious affiliation, donors generally look for those organizations that have both a preferred cause and a reasonable brand reputation. Advertising and fundraising drives provide some other information that can affect which charity a donor chooses. They may also be concerned with scope of impact, from local to international.

These benefits and costs are critical to determining the market outcome of charitable giving. The amount of information available to donors is part of the equation. The quality and quantity of charitable products are typically unknowable to the donors, both before contribution and after. Theoretical literature on asymmetric information, beginning with Akerlof (1970) and used cars, indicates this dearth impedes efficient market functions. If credible proxies for information can be instituted, they can assist the market. There is a large general literature measuring information effects: health codes, audits, third-wave information-based policies, and third-party certification entities such as eco-labeling.

Rather than addressing information, Rose-Ackerman (1982) explicitly models competition between charities through fundraising, but implicitly, she assumes an

⁹ The average household donating in 2004 gave \$2045. However, most of these contributions go to religious groups, with an average of \$1,858. University of Indiana study.

effect of information through donors responding to ‘efficiency’ indicating the cost of charities to raise a dollar. One study explicitly measures the effect of these type of charity ratings (Gordon *et al* 2009) using tax data and the watchdog organization’s rating used in this current study. The researchers measure changes in contributions for a sample of 405 charities for three-year periods between 2002 and 2006. Those authors find a strong positive relationship between higher ratings and contributions in the following year, as intuitive, yet the data limitations therein produce somewhat misleading results.

This current research improves upon this study in many dimensions. First, this research includes all charities rather than a subset and also includes all years since the watchdog group began. This allows better assessment of the effect of ratings and the possibility of changes in effect through the years as ratings gain credibility. Within this broader context, a distinct set of models are applied to determine effects. Second, the data include exact date of publication allowing improved precision and measurement of the timing of ratings. Third, additional historical tax record are appended permitting assessment of the value of ratings as compared to before the ratings are available. And last, the key identification here uses identical methodology of the evaluator, calculating unpublished scores of each charity from the financial data before being rated. This method creates a counterfactual ‘would-have-been’ rating that distinguishes the rating signal of observing the information from the underlying quality.

3 Demand for charity with risk and information

While the donors know the general reputation and goals of the organizations, they frequently do not know the specific actions that charities take toward these goals. Charitable products are typically intangible and the credibility of charities is often uncertain. In addition, there is often a distinction between the donor and the direct beneficiary – the person or group feeling the impact; examples are donations to organizations with projects in homelessness, international aid or the environment, in contrast to charities that provide direct returns to the donor, such

as public radio or art museums. The characteristics and impact of charity work are difficult for the donor to assess. A donor has far less knowledge of the programs and services than the charity, a situation creating a risky investment. This information asymmetry hampers consumer demand and distorts these markets. Furthermore, characteristics of these organizations and their charitable impact are not easily comparable between charities. Ultimately, the charities do not have a clear signal of production for donors to effectively allocate contributions, which may create a suboptimal equilibrium in giving.

The proposed solution is more and better information. As an attempt to alleviate this information asymmetry, charity watchdog groups create ratings based on financial performance of each charity. However, the information must be relevant, understandable, and credible. For verification of these properties, there must also be measurable variation in donor response to the information. The ratings operate directly by increasing donor knowledge and salience of charity efficiency and may shift donations from lower-rated toward higher-rated organizations. The information also could prompt changes indirectly by altering charity response to either negative or positive ratings.

If the rating acts as a signal of charity quality, donors will respond by changing contributions. Setting up a simple model indicates some relevant expected results. The standard charitable giving model is $U_i(x_i, G, g_i)$, where a potential donor has a utility function over a private numeraire good, a public good, and private giving to the public good (Andreoni 1990). However, rather than determine the Nash equilibrium of giving in response to others donations, this research is interested in the effect of information on private giving. Further, much recent literature on giving finds the marginal benefit of giving is private (Vesterlund 2006).¹⁰ Therefore, this setup drops the public good aspect G . This change makes the utility functions independent, allows both dropping the i subscripts and creating a

¹⁰ In the context of internet donations, it is reasonable to assume pure warm glow – there is no conspicuous giving, no reputation gains, and little connection to either the public good or others donations.

form of pure warm glow utility: $U(x + \theta^*g)$.¹¹ A potential donor can choose an allocation between the private good with known returns normalized to 1 per dollar spent and a donation with an uncertain return per dollar spent, θ . The subjective prior probability, determined by the donor, of the return on donations is random variable θ with cumulative distribution $F(\theta)$ and mean μ .¹² With exogenously given wealth ω , we can substitute the budget constraint into the utility, where $U'(\omega) > 0$ and $U''(\omega) < 0$. Let λ denote the fraction donated; then the decision problem becomes

$$\max_{0 \leq \lambda \leq 1} V_F(\lambda) \equiv E_F[U(\omega(\lambda(\theta - 1) + 1))]. \quad (1)$$

Because V is concave, $\lambda^* = 0$ if and only if $V'(0) \leq 0$, and $\lambda^* = 1$ when $V'(1) \geq 0$. Clearly, when the private good returns more than the expected charity gain, $1 \geq \mu$, then $\lambda^* = 0$, no charitable giving. Note also that $\mu > 1$ implies $V'(0) > 0$, therefore a non-trivial allocation of giving charitable is optimal only when these risk-averse donors view the donation as subjectively favorable. The V' is increasing in θ . Also recall that contributions are tax deductible; if donors take this portion as non-risky, a positive donation may result even when the return on the donation is less than unity by the amount of the donor's tax rate, given ω . Furthermore, holding other factors constant, it is reasonable that λ^* increases if and only if F is updated to G , where G is preferred in a stochastic dominance sense.¹³ On the contrary, λ^* decreases when F is preferred to an updated distribution G . In other words, donors who favor giving to charity may learn that returns on donations are better or worse than expected and may change contributions as a result.

The proportional change can be defined as

$$\Delta \equiv \frac{\lambda_G^*}{\lambda_F^*} - 1 \quad (2)$$

¹¹ I am not excited about this functional form, implies perfect substitutes. I am thinking about Cobb-Douglas or just leaving the form implicit.

¹² Given there are many charities, a randomly selected charity has an unknown quality.

¹³ These results have been shown in the analog, demand for risky assets, dating back to Arrow (1965) and as recently as Hollifield & Kraus (2009). This model closely follows the seminal treatment in Gollier (1995) and theorems will not be repeated here.

where λ_F^* gives the former allocation to charity when the donor does not have objective information regarding quality, yet has some subjective probability measure of financial aptitude of charities through other means, such as personal preference, advertisements and reputation. The λ_G^* represents the proportion donated when new information in the form of ratings is presented, and is based on updating beliefs from distribution F to G . The updating, or learning, need not be explicitly specified, but can be conceptualized in a Bayesian framework, wherein the distribution of ratings is incorporated into the priors. Returning to many donors with a fixed and exogenous distribution over wealth, the next step is aggregating the Δ_i 's to the expected average change in donations.

$$\alpha \equiv \frac{\sum_i \Delta_i}{I} \quad (3)$$

Denote this average change in response to information as α , which is the variable of interest measured empirically below, with total population I . This average includes all censored effects, those donors who are not affected by the information and those who do not donate.

This model generates a few hypotheses. Firstly, ratings convey value if and only if they contain credible information the donor did not know, through learning or imparting salience. Therefore Δ will be non-zero given there is sufficient value to the information. Secondly, for two potential donation allocations, higher ratings increase contributions more than lower ratings. This is an artifact of model design, increasing utility in increasing θ . Thirdly, a higher (lower) rating distribution will increase (decrease) the distribution more than lower (higher) rating, given the prior expectation is lower (higher).¹⁴ Lastly, a positive change in Δ indicates the donor had a less favorable view of charitable giving prior to ratings, while a negative change implies an overly auspicious prior view of the charity market. In the empirical section below, the sensibility of this model is further justified with regard to the context of ratings and charitable giving. Given heterogeneity of

¹⁴ These premises abuse precision in notation and should be recast in terms of stochastic dominance.

priors, the results are indicative of average priors. If these are spread, changes will be more difficult to measure.¹⁵

4 Empirical Setting & Evidence

US 990 Tax Returns for Nonprofit Organizations

In the US, a public charity is an organization designated within tax law section 501(c)(3). Although not quite synonymous, this code is often used as a common name for these organizations. Charities are tax-exempt; however, each organization with donation receipts of \$25000 or more must file form 990 annually with the Internal Revenue Service (IRS) regarding revenue and expenses.¹⁶ Nearly 280,000 charities filed in the 2007 tax year. The tax return has more than 100 fields for reporting financial variables for each organization. Revenue variables give an organization's total income, composed of direct public contributions, government grants, payments for program services, and membership dues. Charities list expenses in basic categories such as the costs for programs and services the charity provides, the administrative overhead, and the cost of fundraising. The net of revenue and expenses is the organization's annual excess or deficit. The organizations are non-profit by law, indicating zero excess over the long term. However, charities can maintain assets, investments, and liabilities, which are reported. There are also many descriptive variables such as name, tax ID number, location, fiscal year end date, year of incorporation, year of IRS 501(c)(3) ruling status, and classification code (see appendix for further detail).¹⁷

¹⁵ Both risk-aversion and charitable tax deductions also work against finding an empirical result.

¹⁶ Religious groups, nearly 400,000 in total, are exempt from filing. However, these groups may voluntarily provide the information and about half do. Sourced, March 31, 2010, from <http://nccsdataweb.urban.org/kbfiles/797/Almanac2008publicCharities.pdf>

¹⁷ Organizations are classified by the National Taxonomy of Exempt Entities - Core Codes (NTEE-CC). The hierarchical classification system has 10 primary categories split into 26 major groups (letters of the alphabet) with a finer delineation of these major groups into about 400 total NTEE-CC sectors. For example, a primary category is "Environment and Animals" and 'C' indicates Environmental Groups, which is split into twenty NTEE-CC codes, such as C34: Land Resources Conservation. For comparison purposes, some specifications will be analyzed at the level of these NTEE-CC sectors. (? Maybe – also at level of CN categories, which correspond loosely to the 10 primary categories)

As required by law, these tax returns are available to the public. However, donors tend not to analyze them extensively. First, the volume of charities is extraordinary: in 2005 there were 876,000 public charities – over half of these organizations were required to file with estimated total revenue of \$1.1 trillion for those reporting.¹⁸ The numbers have grown significantly since then as nearly one million charities now are registered. Secondly, the tax forms are complex and confusing. These forms contain hundreds of fields with arbitrary names, difficult-to-assess relationships between the fields, and additional worksheets for revenue and expenses. Furthermore, the forms provide no guidance for identifying ways of comparing values between charities. Two recent reports indicate a continued need to link donors with pertinent, credible easy to understand information about charities.¹⁹

Third-Party Ratings Data

Since April 2002, Charity Navigator (CN), the largest independent charity evaluator, has created measurements of charity performance by synthesizing financial information directly from the tax returns. These metrics are then summarized based on a scale of 0- to 4-stars, a simple index where higher ratings are better, which can assist donors in making informed contributions. Due to budget and time constraints, CN publishes ratings for a subset of all US charities, adding charities in monthly waves. CN is a third-party evaluator, meaning the group is not typically directly involved in donor-charity transactions and does not act on behalf of the rated charities.²⁰ Charities do not pay for ratings. CN is funded as a private foundation, primarily through two individuals and some

¹⁸ <http://nccsdataweb.urban.org/kbfiles/797/Almanac2008publicCharities.pdf>

¹⁹ There is still a need to bridge the information gap: (1) Hewlett Foundation 2009 report on the nonprofit marketplace and, (2) A 2009 report to congress regarding top tax payer problems lists ‘relevant information about needs of nonprofits’ – accessed March 23, 2010.

<http://www2.guidestar.org/rxa/news/articles/2010/nonprofits-make-the-list-of-top-problems-for-the-irs-but-barely.aspx?source=feb10nwltr>

²⁰ Website visitors can link from CN to another site to donate; more discussion below.

advertising.²¹ Also, the charities have no prior knowledge of the first rating. Rather CN aims to serve a broad set of donors, primarily focusing on those organizations with larger total public donations, and selects charities toward achieving this mission.

The selection process is given here and the potential selection bias is addressed below in methods and results. At minimum, organizations must be granted tax-exempt status under US section 501(c)(3) and defined as public charities with four consecutive years of completed IRS 990 forms for CN to determine time-trends. Most large charities meet this last criterion, as any organization will have years of historical growth to gain size and reputation. CN has additional criteria ensuring the nonprofit actively solicits donations and depends on support from individual givers, rather than government grants and foundations. First, CN does not rate charities reporting \$0 in fundraising expenses. Second, the watchdog group prefers charities with public support of more than \$500,000 in the most recent fiscal year. Next, CN excludes charities that receive a large portion of contributions as government grants or charities that have most income from fees charged for programs & services; at least one third of total contributions must come from public support. CN also strives for coverage among charitable causes and across geographic regions, and, as necessary, extends the pool to largest charities in underrepresented classifications. Together, these criteria for a CN-qualified pool of candidates have the implicit effect of narrowing the list to the largest organizations, as measured by total contributions. This subset still covers a large portion of the total annual donations to all US charities.²²

Adding new organizations in waves on a monthly basis, CN gradually expands the number of evaluated charities to create a ‘growing panel’ of data and will

²¹ However, CN recently offered a new option of charity sponsorship, where donors choose a new charity and pay CN’s cost of calculating and publishing the rating. This change occurred after the end of the data set analyzed here.

²² For example for the 2007 tax year, due to heterogeneity in charity sizes, the largest 5% charities by donations received 87% of total US contributions. Need to figure out what percent CN covers.

eventually evaluate all of those charities that remain in the qualified pool.²³ Once selected, CN rates the charity again approximately on an annual cycle, when the latest fiscal year of tax information becomes available. Table 1 gives summary statistics.

TABLE 1: DATA DESCRIPTION				
CHARITY NAVIGATOR SUMMARY STATISTICS				
<i>Rating Observations</i>	30,629			
<i>Charities</i>	5,474			
Variable	Mean	SD	Min	Max
<i>Observations/Charity</i>	5.6		1	9
<i>Fiscal Year (FY)</i>			2000	2008
<i>Publication Year</i>			2002	2009
<i>Publication Lag (mo)</i>	13.3	3.7	3	33
<i>Publication, Quarter in FY</i>	2.3	1.2	1	4
<i>Lag between Pubs (mo)</i>	11.9	3.9	0	65

CN only publishes ratings online: www.charitynavigator.com. The website home page provides general information, related news, as well as an option for user to search for specific charities by name, cause, and other characteristics. Each rated organization has an individual link with summary financial data (see appendix). The most prominent and advertised measure is the *overall rating*, based on a 0- to 4-star scale. The rating is a simple rescaling of *overall score* (ranging from 0 to 70), which is derived by adding ten-point metrics from seven different aspects. The scaling differentiates the charities by efficiency in a systematic manner. *Overall rating* for all charities in all years averages 2.8 stars. One third of the charities receive 3 stars, the median rating, and another 29% are 4-star charities, with only 1.4% getting 0 stars; this distribution is certainly an artifact of scaling but is consistent through the time period.²⁴ The star interpretations, as given by CN from 0- to 4-stars, range over ‘exceptionally poor’, ‘poor’, ‘needs improvement’, ‘good’, and ‘exceptional’. These evaluations relate the acceptable

²³ But CN no longer adds certain types of programs – such as hospitals, higher education, and land trusts.

²⁴ Special scaling considerations are given to some categories, mostly due to differences in overhead between sectors. This will be of note when interpreting regressions and testing sub-groups. There are differences in proportions by category: International groups are well above the average, Environment and Public Benefit are average, while Religion and Arts, Culture, Humanities categories are lower rated, on average.

industry standards and performance of other charities in the same cause. Another linked tab contains two years of the charity’s previous ratings, if the previous years were rated.²⁵ Charities cannot advertise on the CN website, but can put a CN logo on their own website indicating the current rating.

This research measures changes in contributions in response to ratings within a charity, requiring variation in ratings. Looking at the overall historical trend, there is a strong tendency to stay at or near the same rating, as expected due to consistent management and practices within a charity (Table 2). The left column indicates the difference between stars from the previous rating. For example, ‘1’ is a change of one star greater and accessible from all star ratings except if the charity is already at 4-stars. Each column gives the probability of changing stars at the nth rating from the previous rating. The probability of staying at the same star rating, ‘0’, increases through the series, possibly demonstrating long term stability in financial practices.

Star Change	Number of Years Rated								Total	
	2	3	4	5	6	7	8	9	P(chng)	Counts
-4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2
-3	0.003	0.003	0.002	0.002	0.002	0.002	0.000	0.000	0.003	63
-2	0.043	0.042	0.038	0.028	0.021	0.022	0.017	0.014	0.034	857
-1	0.216	0.222	0.195	0.191	0.168	0.159	0.159	0.183	0.197	4,953
0	0.526	0.509	0.520	0.521	0.536	0.568	0.584	0.592	0.528	13,274
1	0.174	0.187	0.204	0.216	0.230	0.218	0.205	0.197	0.200	5,037
2	0.034	0.033	0.038	0.038	0.041	0.027	0.035	0.014	0.035	893
3	0.003	0.004	0.003	0.003	0.002	0.002	0.000	0.000	0.003	72
4	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4
Total	5,280	5,118	4,734	4,074	2,803	2,102	973	71	1	25,155

Table 2

However, the expected probability of a charity-rating changing in any given calendar year is 0.472, indicating adequate variation for testing the effects. Cohort and timing effects may also be present: the 2nd and 3rd ratings are statistically more likely to go down one star than go up one star, while the 6th and 7th ratings are more likely to rise by one than fall by one. While the table does not show which charities are changing, this result indicates a positive trend in ratings over

²⁵ However, this research uses all historical ratings over eight years of data, obtained from Charity Navigator through a confidentiality and data agreement.

many years. Star-changes of more than one are statistically symmetric in both directions over the years. Overall, the table indicates the changes are not random – not as if a die is cast in sequence – and below I address the factors that co-vary with the ratings.

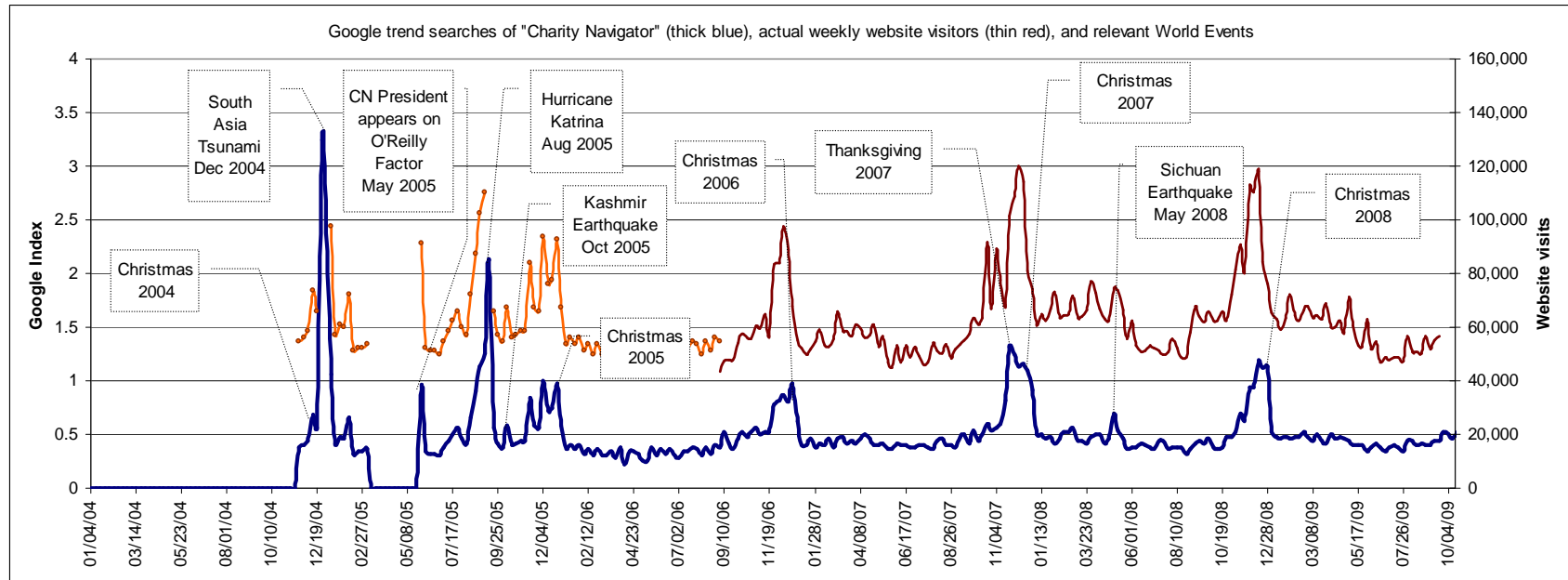
Another concern is reach and readership of these ratings, a signal of how many potential donors that CN impacts. CN has received much media attention since its inception, with interviews on network TV news programs and nearly 80 articles mentioning the group in popular publications of *The New York Times*, *Wall Street Journal*, and *USA Today*.²⁶ Web search behavior and website visits are other indicators of impact (Figure 3). Graphed on the left axis, Google monitors trends in web search and the phrase ‘charity navigator’ shows consistent activity of potential donors. The pattern closely tracks weekly CN website visits, provided by CN, that range from 50,000 to 120,000 unique hits. Huge responses are seen during major world disasters and also at the end of the individual tax year, which coincides with the winter holidays. CN distributes a newsletter at the beginning of each month, eliciting noticeably increased responses. Website visits before September 2006 are unavailable; the figure shows linearly interpolated visits based on the Google Trend index for the earlier time period. Visits are counted daily-to-weekly and may include repeats, making it difficult to assess aggregate audience viewing CN ratings, but the site appears to be known by millions.²⁷

The contributions resulting from online ratings have unique characteristics that justify the above theoretical model. There are no pressure sales or social guilt, we may assume these gifts are motivated by private benefits or desire to increase the charitable goods. Secondly, these are large organizations so contributions of others and overall public good provision do not appear to be factors. Lastly, the ratings affect the contributions reported in tax data only while publicly available.

²⁶ Reference count on lexisnexis.com, January 25, 2010.

²⁷ Charity Navigator self-assessment estimates over 3 million unique website visitors in 2008. Source: <http://www.slideshare.net/CharityNav/valuing-impact-presentation-may-2009?src=embed>, accessed February 2, 1020

FIGURE 1



To respond, the donors must be aware of the information through posting on the website. Therefore ratings for each charity are merged to tax records by corresponding the publish date to the fiscal year affected (see data appendix for more information). Summary of tax data used is given Table 3, for before and after the ratings are available, dollar amounts adjusted to 2007 equivalent.

TABLE 3: TAX DATA DESCRIPTION					
SUMMARY STATISTICS		BEFORE CN		AFTER CN	
<i>Rating Observations</i>		41532		20164	
<i>Charities</i>		5468		4944	
	Variable	Mean	SD	Mean	SD
	<i>Fiscal Year (FY)</i>	2000.0	3.00	2005.5	1.64
<i>1000's of dollars</i>	<i>Total Contributions</i>	11300	50900	23800	92000
	<i>Funraising Expenses</i>	900	3847	1788	6461
	<i>Program Expenses</i>	17300	109000	40000	199000
	<i>Assest</i>	69000	547000	176000	1460000
	<i>Liabilities</i>	12500	96000	43000	571000

5 Effect of ratings on contributions

This study measures effects in two dimensions: (i) distinctions and changes between stars ratings 0 to 4 as derived from an original continuous score, and (ii) before (unobserved) and after (observed) ratings through time. Combinations of these two measures answer various aspects of how much ratings affect donations – the value of receiving each star rating, the difference between ratings, the ability of these scores to represent unobserved heterogeneity of charitable organizations, and the value after observing the information – and determine if these potential changes in contributions are causal.

The following specification captures these aspects along with parameters accounting for other factors that drive annual contributions, *Cont*:

$$Cont_{it} = \alpha^K * Star_{it}^K * Obs_{it} + \rho * Score_{it} + Covars_{it} + v_i + \delta_t + \varepsilon_{it}$$

This regression controls for the different levels of stars through a series of indicators, $Star^K$, where $K = 0, 1, \dots, 4$; and these are our primary variables of interest. The related underlying continuous variable is *Score*, which runs from 0 to

70. The interaction with “*Obs*” indicates the rating is published to the public rather than calculated. The data began in 1995 and *observed* publications of ratings began in 2002.²⁸ The covariates include annual expenses, assets, and liabilities. The series of charity fixed effects, v_i , remove characteristics of each organization that remain constant through the study period. These include founding year, year of IRS ruling the organization as a non-profit, location of operations, general cause and clientele served, scope of services, type of management practices, month of fiscal year end, among others. The δ_t 's are annual intercepts that absorb the average time trend of contributions. Further parameterization will allow distinct time intercepts for each category, such as environmental groups, as trendiness can be distinct in different sectors. The model also assessed separately for each category below.

The ρ gives the marginal slope of an additional unit in *Score* on contributions. *Score* is taken to be the true proxy for quality, a metric designed to index the efficiency of charities. *Star^K*'s, in contrast, are the signals that donors receive, determined at arbitrary breakpoints of the *Score*. Both *Score* and *Star* are calculated for this research for the periods preceding the publication on the rating group's website. However, *Star^K*'s are only included in the model for the *observed* period, acting a visible signal, while *Score* measures the inherent efficiency throughout the entire study. The observed star-rating average is given by $\alpha_K - \alpha_{K-1}$, the difference due to a rating of K-stars relative to another category (below, *Star⁰* is the baseline in some results). Both ρ and the difference between α_K and α_{K-1} are expected to be positive (non-negative) values. Having a higher published score or rating should increase contributions. Otherwise, we expect them to be insignificant. However, a priori, the observation effect is unknown, comparing publication relative to before publication – low star ratings may have a negative effect while high ratings a positive effect; but what is ‘high enough’ is an

²⁸ Charities are added in monthly waves after April 2002, selected across strata of causes (e.g., industry) based on size, historical record, and proportion of public contributions to revenue. This selection process may be correlated with unobserved idiosyncratic panel errors causing bias of coefficients on ratings, which will be addressed in further research.

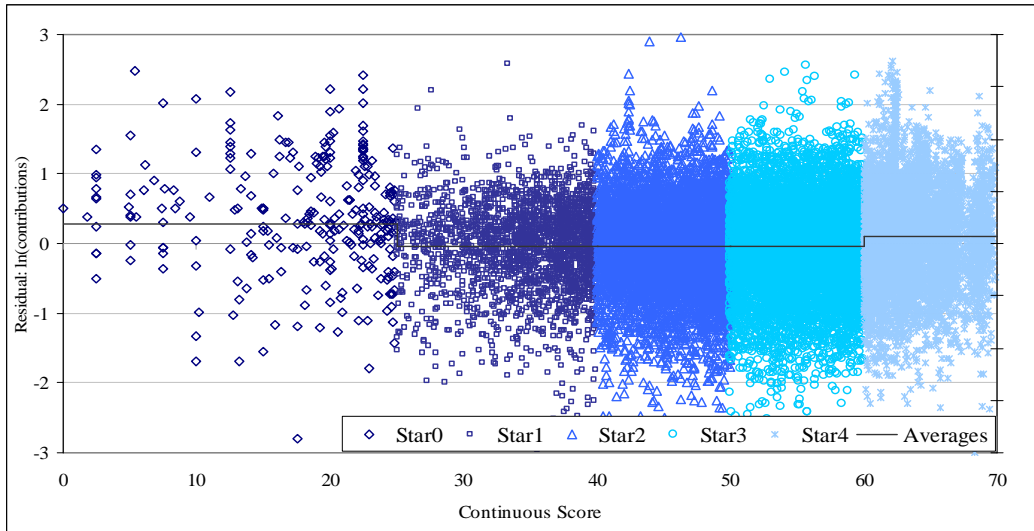
empirical question. The effect of publication may also change through time as the ratings gain esteem. These effects are given by looking at each α_k , which measures the additional average in intercept due to observation of a star rating as compared to not, by controlling for *Score* through all observations.²⁹

Before estimating this model, visually assessment of the residuals from the above specification, less the ratings variables, gives indications of the effects. Also, I take the natural log of the dependent variable and all continuous covariates so that coefficients can be interpreted as percent changes.

$$Resid_{it} = \ln_Cont_{it} - (\widehat{\ln_Covars}_{it} + \hat{v}_i + \hat{\delta}_t)$$

Plotting the residuals over the continuous score variable gives descriptive evidence that contributions respond to the information. For those years of contributions affected by *observed* ratings, the residual variation shows an overall negative trend as score increases, an indication of the sign of ρ , Figure 2. Additionally, average lines for each star-rating produce a distinct stair-step-down pattern, rejecting the hypotheses about $\alpha_k - \alpha_{k-1}$. The line for 4-stars is the exception. Downward trends over the ratings are surprising and discussed below.

FIGURE 2: AVERAGE EFFECTS OF RATINGS ON CONTRIBUTIONS, LN. OBSERVED YEARS.



²⁹ Score is also published on the website, but not highlighted: the media cites only the star-ratings; the CN website uses the stars to rank charities in lists rather than the underlying score; and charities self-advertise only if they receive 4-stars.

To confirm the above figure, results from the model for the *observed* period are given below, Table 4; all models exclude the *unobserved* data. Each charity has multiple ratings at various points in time, thus, the standard errors may be inefficient from a common random effect occurring within a charity. Clustering on charities corrects this. (Unfortunately, both serial correlation, which is likely positive therefore understating SE's, and selection bias may still be issues. I can address the former through tests of autocorrelation and FE-GLS. A Heckman-type selection model might alleviate the latter. Neither has been attempted, yet.)

TABLE 4: ESTIMATED EFFECTS OF RATINGS ON CONTRIBUTIONS, LN. OBSERVED YEARS.

	(1)	(2)	(3)	(4)	(5)	<i>Star-diff</i>	(6)	<i>Star-diff</i>
<i>0-Star, obs</i>		----- omitted			<i>category</i>	-----		
						0.085		0.148
<i>1-Star, obs</i>		0.058	0.087		0.085	(0.048) ⁺	0.148	(0.049) ^{**}
		(0.044)	(0.046) ⁺		(0.048) ⁺	-0.020	(0.049) ^{**}	0.027
<i>2-Star, obs</i>		0.065	0.116		0.065	(0.011) ⁺	0.175	(0.015) ⁺
		(0.045)	(0.050) [*]		(0.048)	-0.011	(0.052) ^{**}	0.033
<i>3-Star, obs</i>		0.072	0.144		0.053	(0.008)	0.208	(0.013) ^{**}
		(0.045)	(0.056) [*]		(0.049)	-0.035	(0.057) ^{**}	0.004
<i>4-Star, obs</i>		0.059	0.148		0.018	(0.009) ^{**}	0.212	(0.011)
		(0.046)	(0.062) [*]		(0.050)		(0.062) ^{**}	
<i>Score, obs</i>	-0.000		-0.002	-0.002			-0.005	
	(0.001)		(0.001) [*]	(0.000) ^{**}			(0.001) ^{**}	
<i>ln_fundraising</i>				0.049	0.049		0.049	
				(0.011) ^{**}	(0.011) ^{**}		(0.011) ^{**}	
<i>ln_prog_exp</i>				0.426	0.424		0.427	
				(0.028) ^{**}	(0.028) ^{**}		(0.028) ^{**}	
<i>ln_assets</i>				0.274	0.274		0.275	
				(0.054) ^{**}	(0.054) ^{**}		(0.054) ^{**}	
<i>ln_liabilities</i>				-0.017	-0.017		-0.017	
				(0.005) ^{**}	(0.005) ^{**}		(0.005) ^{**}	
<i>Observations</i>	20147	20147	20147	20137	20137		20137	
<i>Charities</i>	4944	4944	4944	4944	4944		4944	
<i>R-squared</i>	0.02	0.02	0.02	0.19	0.19		0.19	

Robust SEs in parentheses; ⁺, ^{*}, and ^{**} significant at 10%, 5%, and 1%, respectively.
Each model includes charity and annual fixed effects.

The model improves on the figure by controlling for all covariates simultaneously, giving average marginal effects on within-charity variation of contributions. Column (1) uses only *Score* as the indicator of quality and measures the average slope. The second results, (2), include *observed* ratings of stars and removes *Score*, while (3) has both star-rating and *Score*. Models (4) through (6) repeat the previous models in the same order, but include the covariates. This exercise illustrates the relationship of the covariates, expenses

and assets, with the ratings. Table 4 also reports the differences between star-ratings and related significance for specifications (5) and (6).

Because of the log-linear model form and indicator variables, the *star* effects are approximate percent changes.³⁰ These results confirm some of our intuition from above, but also offer surprises. The differences between *observed* star ratings, $\alpha_K - \alpha_{K-1}$, are positive (non-negative) values when underlying Score is also present (3) and (6). Having a higher published rating increases contributions: observing 3- rather than 2-stars gives a 3.3% increase in contributions, while 2- to 1-stars is 2.7%, and so on. The ρ on *observed Score* is predicted to be positive, however, these models indicate a small negative slope – within each rating or without the other controls; a higher score decreases contributions slightly. This is also true of the star-rating when the continuous score is omitted – higher ratings, lower scores. One explanation is that higher score is correlated with lower fundraising and thus lower donations; however, the negative relationship remains with and without including the fundraising covariate. More understanding of this result is needed.

To distinguish between *observed* and *unobserved*, all the residuals are plotted through time, Figure 3. Because charities are introduced to the ratings at different times, the data must be aligned relative to the first year rated, graphing the coincident timing of publication. The x-axis denotes the n^{th} rating, where 0^{th} , by definition, is the year before the charity's 1st inclusion. The left trend line shows that average residual variation increases over subsequent observations as the introduction of ratings approaches. However, the publication of ratings appears to cause an abrupt decrease in contributions. The right-area slope then continues at about the same increase as the pre-publication trend. If the graph only includes 7-9 years of pre-rating data, as symmetric with the post-rating data, the slope for years prior to introduction of ratings is steeper and the downward jump larger.

³⁰ None of the reported percents are corrected for the log-linear/dummy-variable bias: $\exp^{\hat{\beta}} - 1$.

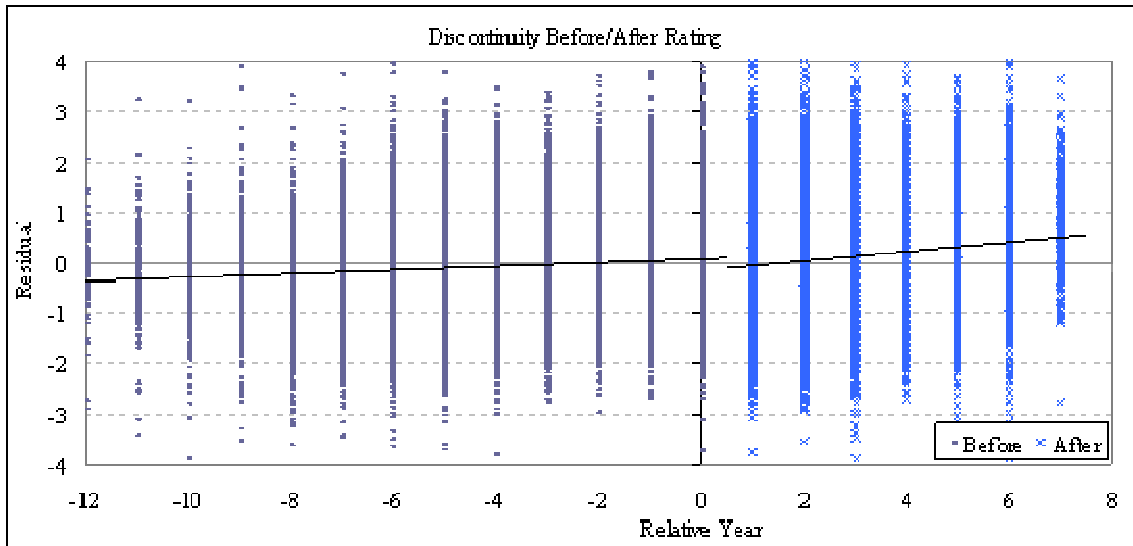


FIGURE 3: AVERAGE TREND OF RATINGS EFFECTS ON CONTRIBUTIONS, LN. BEFORE & AFTER.

The full model estimates each of these aspects, ratings before and after, at the same time while holding the others constant. Table 5, below, shows the full results with specifications mirroring the previous analyses in Table 4; clustered standard errors are reported. Column (1) uses *Score* for the entire time, calculated for period prior to ratings; (2) includes *observed* ratings of stars and removes *Score*, while (3) has both star-rating and *Score* giving differences between stars ratings for observed period, and slope of the continuous score, for both before and after publication of ratings through time. Again, models (4) through (6) repeat the previous models in the same order, but include the covariates. Table 5 also reports the differences between star-ratings and related significance for specifications (5) and (6).

In all cases, estimated effects of ratings are interpreted as weighted averages of the effects of *observed* ratings in all years of publication and this is the signal value. These are the coefficients on *observed* ratings when *unobserved* data is included in the model for reference. Contrary to the expectation of information, all ratings have a negative effect, but less so as the ratings improve from 0-through 4-stars. In specifications (2) and (5), where only dummies control for ratings, having a published 4-star rating cannot be differentiated from being

unpublished. In model (6), inclusive of the continuous *Score*, having a published 4-star rating is 2.4% worse than being unpublished.

TABLE 5: ESTIMATED EFFECTS OF RATINGS ON CONTRIBUTIONS, LN. ALL YEARS.

	(1)	(2)	(3)	(4)	(5)	<i>Star-diff</i>	(6)	<i>Star-diff</i>
<i>0-Star, obs</i>		-0.443 (0.059)**	-0.305 (0.058)**		-0.270 (0.046)**	} 0.060 (0.046) } 0.052 (0.013)** } 0.064 (0.009)** } 0.082 (0.010)**	-0.236 (0.046)**	} 0.050 (0.046) } 0.040 (0.014)* } 0.051 (0.009)** } .07077 (0.010)**
<i>1-Star, obs</i>		-0.409 (0.018)**	-0.306 (0.018)**		-0.211 (0.016)**			
<i>2-Star, obs</i>		-0.282 (0.014)**	-0.229 (0.013)**		-0.158 (0.012)**			
<i>3-Star, obs</i>		-0.165 (0.012)**	-0.164 (0.012)**		-0.094 (0.010)**			
<i>4-Star, obs</i>		-0.006 (0.014)	-0.050 (0.014)**		-0.013 (0.011)			
<i>Score, obs</i>	0.008 (0.000)**		0.007 (0.000)**	0.003 (0.000)**			0.002 (0.000)**	
<i>ln_fundraising</i>				0.035 (0.003)**	0.034 (0.003)**		0.034 (0.003)**	
<i>ln_prog_exp</i>				0.043 (0.003)**	0.046 (0.003)**		0.044 (0.003)**	
<i>ln_assets</i>				0.363 (0.023)**	0.361 (0.023)**		0.358 (0.023)**	
<i>ln_liabilities</i>				0.010 (0.003)**	0.010 (0.003)**		0.010 (0.003)**	
<i>Observations</i>	61604	61604	61604	61561	61561		61561	
<i>Charities</i>	5468	5468	5468	5468	5468		5468	
<i>R-squared</i>	0.19	0.18	0.20	0.36	0.37		0.37	

Robust SEs in parentheses; +, *, and ** significant at 10%, 5%, and 1%, respectively.
Each model includes charity and annual fixed effects.

As a falsification test of the theory that stars are signals – effective only when published – I can include calculated, but unpublished stars and determine if these ratings determine contributions when unseen. These effects, weighted averages of the effects of *unobserved* ratings prior to publication, are statistically insignificant, and are not reported. With score controlling for efficiency qualities captured by the rating design, the unpublished stars apparently do not capture charity characteristics intrinsically know to donors that affect contributions.

Changes in ratings, rather than levels, are also analyzed by defining the absolute change relative to the previous rating, such as a move from 3- to 4-stars. This gives 25 indicator variables labeled *rating_{AtoB}*, where *A* is the previous rating and *B* is the current.³¹ This approach has the added advantage of using charities that

³¹ *A-to-B* combinations are rarer as the difference gets larger, such as 1-to-4, and therefore not all 25 coefficients are statistically viable, as seen in results.

remain at the same ratings. To determine the effects of observing a change, $rating_{AtoB}$ is interacted with Obs . Using the full specification from above, Table 6 gives the coefficients on $rating_{AtoB} * Obs$: the percent changes in contributions due to an observed change in stars from the previous year relative to the pooled unobserved data. The coefficients should increase toward the right along rows or down-right on a diagonal, and for the most part they do. Again, all significant coefficients are negative, indicating a decrease in contributions due to publication of ratings. If included, none of the *unobserved* star-changes in ratings are significantly different than zero (model not reported, but acts as a falsification test).

TABLE 6: ESTIMATED EFFECTS OF CHANGES IN RATINGS ON CONTRIBUTIONS, LN.

$n-1^{th} \setminus n^{th}$	0-stars	1-stars	2-stars	3-stars	4-stars
0-stars	-0.374 (0.087)**	-0.126 (0.094)	-0.165 (0.122)	-0.241 (0.183)	-0.458 (0.334)
1-stars	-0.215 (0.050)**	-0.175 (0.022)**	-0.165 (0.023)**	-0.055 (0.033)*	-0.251 (0.154)
2-stars	-0.143 (0.122)	-0.167 (0.023)**	-0.136 (0.017)**	-0.086 (0.017)**	-0.094 (0.029)**
3-stars	0.118 (0.187)	-0.131 (0.034)**	-0.119 (0.015)**	-0.081 (0.013)**	-0.035 (0.016)*
4-stars	-0.360 (0.041)**	-0.043 (0.121)	-0.059 (0.026)*	-0.050 (0.014)**	0.021 (0.013)

Observations: 54624, Charities: 5459, R-squared: 0.30

Robust standard errors in parentheses; *, and ** significant at 5%, and 1%, respectively

The model includes the covariates from specification above and Score (with a coefficient of 0.002**), charity and annual fixed effects. Unobserved periods compose the comparison groups.

The effect of having observed low scores (0- or 1-stars) two years in a row is quite costly, with a 17.5% to 37.4% decrease in contributions. Unfortunately, even moving from a 3- to a 4-star rating reduces contributions relative to before publication. One would expect that a star increase would be better than a star decrease, but the two seem the same.

Again using the preferred specification from above, Table 7 gives the coefficients by category, the percent changes in contributions due to an observed star rating from relative to the pooled unobserved data. The categories are based on classification codes from the tax forms, the NTEE codes, and represent broad

sectors. While each category has a general aim, charities within the sectors can range over many types such advocacy or action groups. The results of estimation show higher star-ratings are better. Yet even within category, most coefficients on stars remain negative and significant, indicating a decrease in contributions due to publication of ratings. However, a few are positive, indicated in bold – religion and health sectors gain increased contributions by the observation of higher ratings. Still, these categories along with education and human services are hardest hit by 0-star ratings.

TABLE 7: ESTIMATED EFFECTS OF RATINGS, BY CATEGORY ON CONTRIBUTIONS, LN.

	<i>Animals</i>	<i>Art, Culture, Humanities</i>	<i>Education</i>	<i>Envrnmnt</i>	<i>Health</i>	<i>Human Services</i>	<i>Internat'l</i>	<i>Public Benefit</i>	<i>Religion</i>
<i>0-Star, obs</i>	-0.143 (0.123)	-0.163 (0.104)	-0.258 (0.123)*	-0.279 (0.250)	-0.256 (0.104)*	-0.327 (0.097)**	-0.120 (0.184)	-0.181 (0.117)	-0.284 (0.114)*
<i>1-Star, obs</i>	-0.224 (0.063)**	-0.140 (0.043)**	-0.173 (0.050)**	-0.235 (0.058)**	-0.150 (0.044)**	-0.149 (0.039)**	-0.249 (0.059)**	-0.177 (0.037)**	-0.175 (0.042)**
<i>2-Star, obs</i>	-0.190 (0.054)**	-0.133 (0.028)**	-0.150 (0.038)**	-0.150 (0.048)**	-0.105 (0.034)**	-0.132 (0.027)**	-0.199 (0.045)**	-0.132 (0.028)**	-0.046 (0.032)
<i>3-Star, obs</i>	-0.128 (0.042)**	-0.073 (0.027)**	-0.123 (0.030)**	-0.116 (0.040)**	-0.033 (0.028)	-0.112 (0.024)**	-0.095 (0.044)*	-0.084 (0.025)**	0.047 (0.028)+
<i>4-Star, obs</i>	-0.064 (0.043)	-0.092 (0.034)**	-0.054 (0.033)+	-0.055 (0.038)	0.066 (0.030)*	-0.035 (0.024)	0.026 (0.046)	-0.072 (0.030)*	0.138 (0.043)**
<i>Score, obs</i>	-0.000 (0.001)	-0.000 (0.001)	0.003 (0.001)**	-0.000 (0.001)	0.002 (0.001)*	0.003 (0.001)**	0.001 (0.001)	0.002 (0.001)**	0.002 (0.001)*
<i>ln_fund</i>	0.044 (0.008)**	0.028 (0.007)**	0.033 (0.007)**	0.029 (0.009)**	0.047 (0.010)**	0.035 (0.005)**	0.040 (0.010)**	0.027 (0.006)**	0.017 (0.005)**
<i>ln_prog_exp</i>	0.034 (0.008)**	-0.002 (0.006)	0.036 (0.010)**	0.029 (0.008)**	0.043 (0.008)**	0.073 (0.008)**	0.064 (0.010)**	0.040 (0.008)**	0.054 (0.011)**
<i>ln_assets</i>	0.323 (0.123)**	0.408 (0.071)**	0.310 (0.086)**	0.471 (0.027)**	0.360 (0.061)**	0.373 (0.021)**	0.327 (0.041)**	0.389 (0.046)**	0.230 (0.053)**
<i>ln_liabilities</i>	-0.010 (0.006)+	0.005 (0.010)	0.023 (0.009)*	0.019 (0.008)*	0.023 (0.007)**	0.005 (0.006)	0.014 (0.005)**	0.011 (0.011)	0.004 (0.006)
<i>Observations</i>	4539	8072	6519	3745	6656	12858	5300	9877	3995
<i>Charities</i>	393	676	563	337	583	1150	514	893	359
<i>R-squared</i>	0.34	0.27	0.30	0.40	0.44	0.45	0.48	0.37	0.40

Robust SEs in parentheses; +, *, and ** significant at 10%, 5%, and 1%, respectively.

Each model includes charity and annual fixed effects.

Preliminary work on size of charity indicates larger negative results for smaller charities, but all results are consistent with the overall findings. Results for only the Star^K coefficients are reported in Table 8, where size tertiles are distinguished using a charity's average contributions over the entire time period. The first tertile includes the smaller charities, second is medium, and third is largest. The percent changes for smallest charities are typically largest in each star rating, except for 0-

stars. In the 4-star category, the small charities are the only one with a strongly significant negative result.

TABLE 8: ESTIMATED EFFECTS OF RATINGS, BY SIZE ON CONTRIBUTIONS, LN.

	<i>Small</i>		<i>Medium</i>		<i>Large</i>	
<i>0-Star, obs</i>	-0.251 (0.070)**	0.040**	-0.308 (0.070)**	0.113**	-0.162 (0.121)	-0.008
<i>1-Star, obs</i>	-0.211 (0.026)**	0.002	-0.195 (0.028)**	0.048**	-0.170 (0.033)**	0.058**
<i>2-Star, obs</i>	-0.209 (0.023)**	0.083**	-0.147 (0.019)**	0.065**	-0.112 (0.026)**	0.009*
<i>3-Star, obs</i>	-0.126 (.019)**	0.076**	-0.082 (0.018)**	0.053**	-0.103 (0.024)**	0.064**
<i>4-Star, obs</i>	-0.050 (0.022)*		-0.029 (0.021)		-0.039 (0.023) ⁺	

Robust SEs in parentheses; ⁺, *, and ** significant at 10%, 5%, and 1%, respectively.
Each model includes all covariates, and charity and annual fixed effects.

Finally, an additional empirical model use regression discontinuity, which simply looks at those charities near the star-rating breakpoints. Analogous to grades in a classroom where an 89% is a 'B' yet a 90% is an 'A', charities with similar scores but a one-star difference in ratings can have distinct outcomes in contributions. In this model, the differences between stars ratings are smaller, and there is no difference between 0- and 1-stars. However, the jump to 2-stars is 2%. Then the step up to 3-stars is 4%, and an additional discontinuity of 4% at 4-stars. These findings are consistent with the average marginal value results in Table 5.

6 Discussion

This research gives strong evidence that donors respond to these rankings. As expected, higher stars give higher contributions relative to lower stars. However, there are a few surprising results. First, contributions do not appear as related to the efficiency performance of the charity, measured by score, as much as the signal of the star rating: when only the observed data are analyzed the organizations that rank highest in efficiency by score receive a smaller share of donations in those periods. Second, when comparing contributions before and after information is available, the ratings seem to 'raise the bar' of scrutiny – most

charities will see donations decline, no matter the stars earned. Furthermore, the research shows that the relationship between donations and changes in performance is identical if a group goes up one star or down one star, indicating donors may update their prior expectations with average ratings instead of looking at only the most recent.

Accordingly, the overall donations to the rated charities fall after being rated. We might wonder where this money goes. One indication comes from the ability to “donate now” at the CN website. The link redirects donors to an aggregate payment site, called *Network for Good*. This site allows donations using credit cards and payment for multiple charities simultaneously. However, there is an implicit agreement with the respective charity to which the donations are given; *Network for Good* receives three percent of each donation. This cut accounts almost perfectly for the loss in 4-star charities. Therefore, CN could assist by directing donors to the specific charity to give rather than *Network for Good*. Additionally, aggregate donations (including unrated charities) do not seem to fall during same time period. Donors maybe substituting donations to non-rated charities from rated charities. As this paper develops further, the research will consider contribution shifting between rated and unrated charities. For future work, the time-series of aggregate US contributions will be used to test the overall economic impact of charity evaluation. These results demonstrate potential unintended consequences of shifting donations based on efficiency rather than effectiveness. The primary finding is that donors over-evaluate charities, relative to these ratings, and may not want to know.

Appendix: Data Composition and Calculations

This section explains the collection, composition, and calculations of the data. The core data are from the largest US charity evaluator, Charity Navigator (CN), which creates ratings by synthesizing many financial variables directly from the tax returns. These ratings are summarized into a scale of 0- to 4-stars, a simple index where higher ratings are better, which can assist donors in making informed contributions. I classify this baseline set of variables into three categories, explained further below: IRS 990 financial reported data, CN calculations, and CN ratings. Each is explained in more detail in the sections that follow. The first group indicates that charities report the data annually on tax forms; these are historical and longitudinal financials used to create the ratings. The second class denotes intermediary performance criteria CN chooses to target for the ratings, including the seven metrics that funnel into the ratings. And the final category

Form 990	Return of Organization Exempt From Income Tax		OMB No. 1545-0047
Department of the Treasury Internal Revenue Service	Under section 501(c), 527, or 4947(a)(1) of the Internal Revenue Code (except black lung benefit trust or private foundation)		2008 Open to Public Inspection
▶ The organization may have to use a copy of this return to satisfy state reporting requirements.			
A For the 2008 calendar year, or tax year beginning		, 2008, and ending	
B Check if applicable: <input type="checkbox"/> Address change <input type="checkbox"/> Name change <input type="checkbox"/> Initial return <input type="checkbox"/> Termination <input type="checkbox"/> Amended return <input type="checkbox"/> Application pending		C Name of organization Doing Business As Number and street (or P.O. box, if mail is not delivered to street address) Room/suite City or town, state or country, and ZIP + 4	D Employer identification number : E Telephone number ()
I Tax-exempt status: <input type="checkbox"/> 501(c) () (insert no.) <input type="checkbox"/> 4947(a)(1) or <input type="checkbox"/> 527		F Name and address of principal officer: H(a) Is this a group return for affiliates? <input type="checkbox"/> Yes <input type="checkbox"/> No H(b) Are all affiliates included? <input type="checkbox"/> Yes <input type="checkbox"/> No If "No," attach a list. (see instructions) H(c) Group exemption number ▶	
J Website: ▶		G Gross receipts \$	
K Type of organization: <input type="checkbox"/> Corporation <input type="checkbox"/> Trust <input type="checkbox"/> Association <input type="checkbox"/> Other ▶		L Year of formation: M State of legal domicile:	
Part I Summary			
Activities & Governance	1 Briefly describe the organization's mission or most significant activities:		
	2 Check this box <input type="checkbox"/> if the organization discontinued its operations or disposed of more than 25% of its assets.		
	3 Number of voting members of the governing body (Part VI, line 1a)	3	
	4 Number of independent voting members of the governing body (Part VI, line 1b)	4	
	5 Total number of employees (Part V, line 2a)	5	
	6 Total number of volunteers (estimate if necessary)	6	
	7a Total gross unrelated business revenue from Part VIII, line 12, column (C)	7a	
b Net unrelated business taxable income from Form 990-T, line 34	7b		
Revenue	8 Contributions and grants (Part VIII, line 1h)	Prior Year	Current Year
	9 Program service revenue (Part VIII, line 2g)		
	10 Investment income (Part VIII, column (A), lines 3, 4, and 7d)		
	11 Other revenue (Part VIII, column (A), lines 5, 6d, 8c, 9c, 10c, and 11e)		
	12 Total revenue—add lines 8 through 11 (must equal Part VIII column (A) line 12)		

Figure A1: IRS Form

describes the three ratings, split into efficiency and capacity scores, then summarized by overall score.

US IRS 990 Form, Tax Data

These variables are fields from the tax form pictured above, digitized and named accordingly.

charity name: often has variations from year to year (e.g. Fnd versus Foundation)

ein: unique US tax/employee identification number

fye: fiscal year end of the associated financial data

total contributions: public and government grant support

program service revenue: sales from goods or services

membership dues: income from club membership, distinguished from donations

other revenue: income other than contributions, dues, and program revenue

total revenue: revenue from all sources

program expenses: costs of programs/services

administrative expenses: operational costs

fundraising expenses: costs of raising funds and capital

payments to affiliates: transfers to associated organizations

total expenses: total of the four above costs

total assets: balance sheet of valuable cash and capital

total liabilities: balance sheet of owed claims against organization

CN Calculations

These variables, performance measures targeted by CN, are calculated from the above data. Charity Navigator provides descriptions of all calculations as a transparency policy. All calculations and scales are described in detail at <http://www.charitynavigator.org/index.cfm/bay/content.view/catid/2/cpid/48.htm> (accessed September 15, 2009). The seven performance metrics directly used in ratings are marked: efficiency measures are underlined, capacity measures double underlined.

orgid: CN identification number

category: CN designations of the broad area of charitable activity, related loosely to IRS NTEE-CC classifications

cause: CN designations, sub-classification of categories.

state: US state location of charity, does not indicate region of impact

scope: Region of impact; local/regional, national, international

publish date: the calendar date information made available to public

functional expenses: sum of program, administrative, and fundraising expenses

program expense ratio: proportion of programmatic costs to total functional expenses

administrative expense ratio: proportion of administrative costs to total functional expenses

fundraising expense ratio: proportion of fundraising costs to total functional expenses

fundraising efficiency: ratio of fundraising costs to functional expenses, interpreted as the average cost to raise a dollar

primary revenue: sum of three fields from the 990 form (total contributions, program service revenue, and membership dues)

primary revenue growth: average annual change in primary revenue over T years

program expenses growth: average annual change in program expenses over T years

working capital: current net liquid assets, sum of six assets fields from the 990 form (cash, savings, accounts receivable, pledges, grant, investments) less three liabilities fields (accounts payable, grants payable, other)

working capital ratio: estimates how many years working capital can sustain the current total expenses

CN Scores and Ratings

The most prominent, advertised measure, and the focus of this analysis, is the *overall rating*, based on a 0- to 4-star scale. The rating is a simple rescaling of *overall score* (ranging from 0 to 70), which is derived by adding ten-point metrics from seven different aspects. (See schematic, Figure A1). As noted above, the calculations group into ‘expense data’ and ‘trend data.’ The four measures of expenses are each converted to a ten point scale and added to determine the *efficiency score* (0 to 40 points). Charity Navigator then converts the numeric scores to a 0- to 4-star system, the *efficiency rating*. The three measures of ‘trend data’ calculate growth and asset capacity of an organization; again, they are each converted to a ten point scale and added to determine the *capacity score* (0 to 30

points) the numeric scores are converted to 0 to 4 stars, the *capacity rating*.³² The *overall rating*, also 0- to 4-star values, is simply a scaling at given thresholds of the *overall score*, which is the sum of capacity and efficiency ratings, ranging

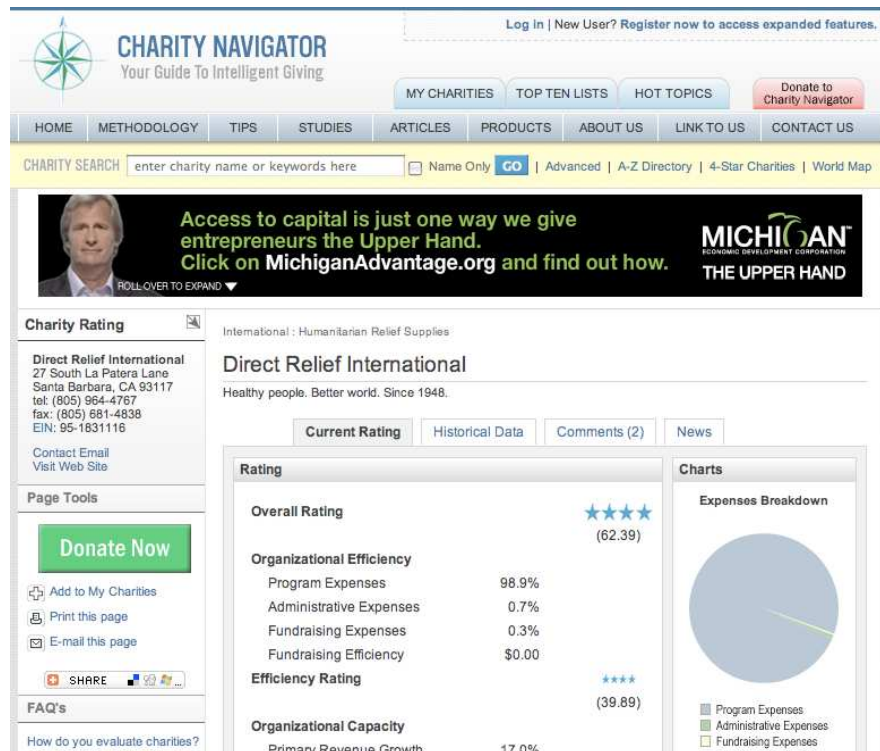
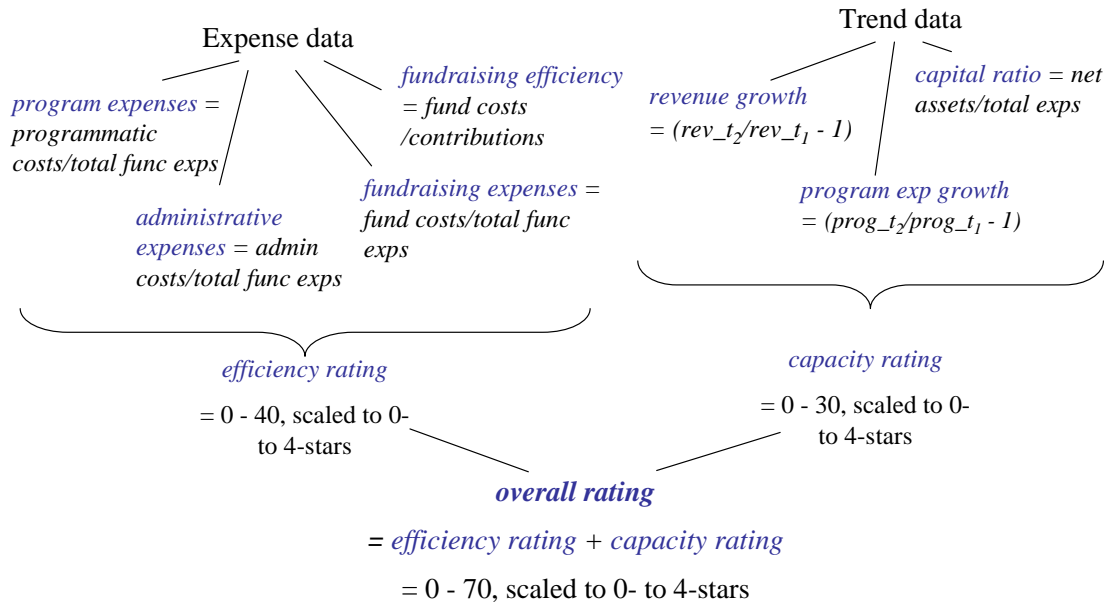


Figure A2: Website screen capture of a charity, Direct Relief International; their most recent *overall rating* is 4-stars.

³² Co-linearity of expenses and capacity may occur and should be addressed.

from 0 to 70 raw score. The threshold ranges are arbitrarily assigned and constant through the study period: Star0 = 0-24.9, Star1 = 25-39.9, Star2 = 40-49.9, Star3 = 50-59.9, Star4 = 60-70.

Additional Data

To create ratings prior to the initiation of Charity Navigator, further historical data are necessary. No single alternative source provides both the ideal time length and the needed breadth of variables. Digital filing has only recently been available, making paper filing the source of data and creating a need to digitize after the fact. The National Center for Charitable Statistics (NCCS), a project of the Urban Institute partially digitizes the annual 990 form submitted by each charity into Core Files – meaning not all variables from the written form are available in these data sets. Particularly, while decades of data are offered for purchase, unfortunately, not all the necessary components for the CN ratings are included in this source. The pertinent variables available from the NCCS Core Files are *charity name, ein, fye, total contributions, program service revenue, membership dues, other revenue, total revenue, fundraising expenses, total expenses, total assets, and total liabilities*. To complete the calculations, three further expense variables are needed (*program expenses, administrative expenses, payments to affiliates*), as well as many variables for determining working capital (*cash, savings, accounts receivable, pledges, grant, investments, accounts payable, grants payable, other liabilities*). These NCCS Core Files are designated by the year the organization filed and currently run through 2007. This research uses available files since 1993.

In an attempt to supplement the NCCS Core Files, I assemble data from two other sources. Guidestar.org is a nonprofit web-service that gathers and provides information about charities to donors and researchers. Academic access allows viewing of digital copies of 990 forms and additional digitized variables. Combined with the Core Files, these variables include all data for creating CN metrics, but the Guidestar data for each charity start in 1997 or later.

I also append IRS Statistics of Income files, available from 1992, a subset in returns produced by the department to analyze elements of the tax system. These data have all the variables necessary for rating calculations, filling in historical variables missing from the other sets, but are small random samples from each year of the complete population of filed charity returns. The missing data problem is improved, yet remains. Most CN charities are not in the file in any given year. This sampling leaves regression results unbiased, but affects regression efficiency.

Unpublished Ratings and Interpolating missing data

The pivotal estimation compares ‘would-have-been’ historical ratings to published ratings. Creating historical unpublished ratings follows CN ratings calculations exactly when all the essential variables are available. As a validation of correct calculations, I first attempt to replicate the known ratings, which I can do with R^2 of 0.999 and MSE of 1.2. The supplementary tax data provide a historical trend of unpublished ratings for most charities. However, there are many observations for which most of the historical data are available but not each of the metrics necessary for the calculation. Because a longer time trend will better comparison of unpublished verses published ratings (and will help alleviate unobserved variable bias), it may be useful to interpolate the missing values. I then calculate further unpublished ratings from these interpolated variables. Results excluding these interpolated ratings are available upon request.

Figure A3 compares the distribution for all charities of unpublished, calculated overall-scores including the interpolated scores (left panel) for the historical data to the published overall-scores (right panel). The distributions are very similar, lending credibility to both method of calculation and interpolation. The means and medians are nearly identical, while the spread of the calculated score is slightly larger. The two distributions need not be the same because the underlying data represent two distinct time periods, pre-CN and present-CN periods.

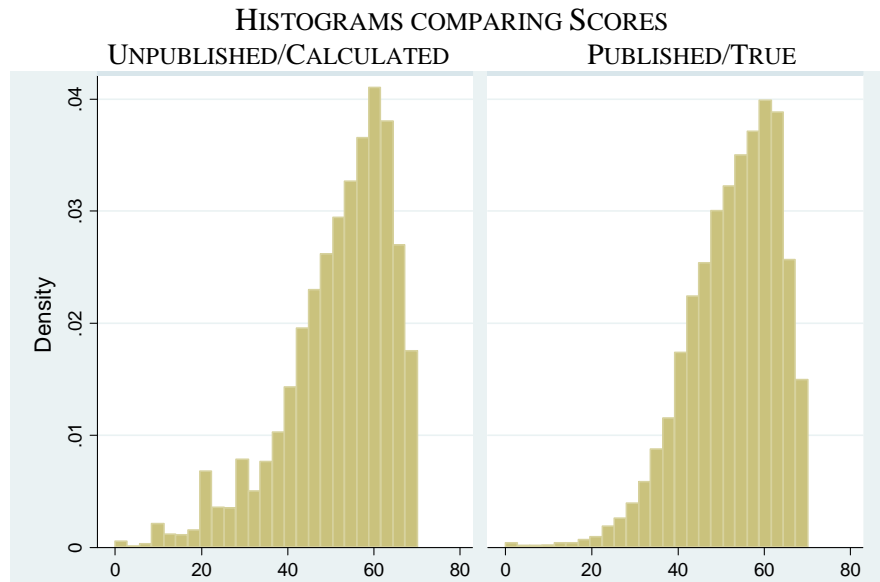


Figure A3

Despite missing data, interpolations are reasonable for many reasons. When variables are missing from various sources, they are typically the fields related to *working capital* and the components of *expenses*. Secondly, I can assume ‘missing’ occurs at random, due to some values not being digitized for many charities in the source dataset while other contemporaneous variables are digitized. This allows unbiased interpolation. Two naïve interpolations fill in values for a charity using simple within-averages and time-moving averages of available data. While both these approaches reduce the variation by restricting predictions to the known value range, the latter accommodates a time trend. A more sophisticated approach models and predicts using regressions of related, contemporaneous variables. I created an indicator of the fraction of the rating components out of the twelve interpolated variables, ranging from 0 to 1. The median interpolation weight is 0.08 and the average is 0.18.³³

The interpolations leave coefficients unbiased but can create inconsistent standard errors. If interpolation is created from variation within each group, I can use fixed

³³ Double check these values.

effect GLS on the charity panels to create an appropriate variance structure. If the entire dataset is used to construct missing data, I follow Shao and Sitter (1996) to correct the standard errors: bootstrap on charities, impute data and run the regression of interest within the bootstrap, use 200 repetitions, and output corrected standard errors.

Merging the ratings to the affected fiscal year

Published ratings are merged to tax records by using the employer identification number, which uniquely identifies each organization, and then aligning the publish date with the fiscal year affected for each charity. Figure A3 gives an example for one organization, demonstrating that ratings are calculations from past financial records (the year in parentheses) and affect contributions during the publication period, which can vary in length. This alignment must be reconciled for appropriate results.

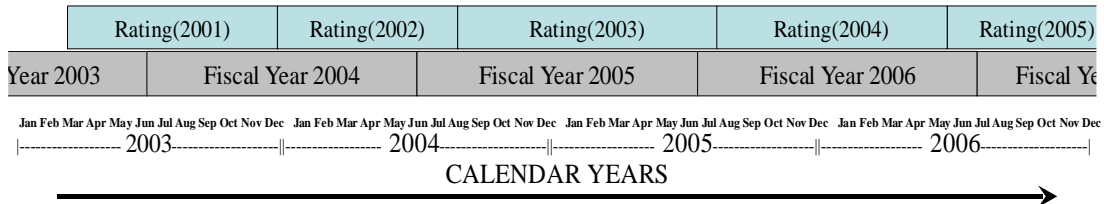


Figure A3

Changes in fiscal year

Unlike individuals who have an April 15th tax deadline, each charity can choose the end of its fiscal year (FYE). For the majority of organizations in the data set the FYE does not change – the accounting period ends on same date each year. However, 802 charities (14.6%) do change at least once during the 16 year study period. If management changes the FYE, the charity must file a short-year return for the interim months. These changes are explicit in the data set shown as changes in the month filed, unfortunately, the short-year data are unavailable for many years, creating difficulty for assessment including misspecification of the fixed effect. The changes are not made in anticipation or response to ratings, but rather due to changes in management; therefore, I keep the data when it remains

unchanged. When the change occurs within the CN period, analysis includes the charity's data prior to the change. When the change occurs prior to CN existence, analysis includes the charity's data after the change. Together, adjusting causes loss of 5.2% of the observations. If these 802 charities are dropped altogether, 15.5% of the data is lost. Dropping data with fiscal year changes should not bias the results.

References

- Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*. 84(3): 488 – 500.
- Andreoni, James. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy* 97:1447 – 1458.
- _____. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving" *Economic Journal* 100:464 – 77.
- Arrow, Kenneth. 1965. "Aspects of the Theory of Risk-Bearing," Lecture in Helsinki: Yrj Hahnsson Foundation.
- Auten, Gerald, Holger Sieg, and Charles Clotfelter. 2002. "Charitable Giving, Income, and Taxes: An Analysis of Panel Data." *American Economic Review* 92:371 – 82.
- Bergstrom, Ted, Lawrence Blume, and Hal Varian. 1986. "On the Private Provision of Public Goods," *Journal of Public Economics* 29(1):25-49.
- DellaVigna, Stefano, John A. List, and Ulrike Malmendier. December 2009. "Testing for Altruism and Social Pressure In Charitable Giving." NBER Working Paper 15629.
- Gollier, Christian. 1995. The Comparative Statics of Changes in Risk Revisited. *Journal of Economic Theory*. 66(2): 522 –535.
- Gordon, Teresa P., Cathryn L. Knock, and Daniel Neely. 2009. "The Role of Rating Agencies in the Market for Charitable Contributions: An Empirical Test," *Journal of Accounting and Public Policy*, .
- Hollifield, Burton, and Alan Kraus. 2009. "Defining Bad News: Changes in Return Distributions That Decrease Risky Asset Demand," *Management science*. 55(7): 1227 –1236.
- Payne, Abigail. 1998. "Does the government crowd-out private donations? New evidence from a sample of non-profit firms" *Journal of Public Economics* 69:323-45.
- Prince, Russ and Karen File. 1994. *The Seven Faces of Philanthropy : A New Approach to Cultivating Major Donors*. San Francisco. The Jossey-Bass Nonprofit Sector Series.
- Randolph, William. 1995. "Dynamic Income, Progressive Taxes, and the Timing of Charitable Contributions." *Journal of Political Economy* 103:709 – 38.

Rose-Ackerman, Susan. 1982. "Charitable Giving and 'Excessive' Fundraising." *Quarterly Journal of Economics* 96:193 – 212.

Shao, Jun and Randy R. Sitter. 1996. "Bootstrap for Imputed Survey Data." *Journal of the American Statistical Association* 91(435): 1278 – 1288.

Tullock, Gordon. 1966. "Information without Profit." pp. 141 – 59 in *Papers on Non-Market Decision Making*, edited by Gordon Tullock. Thomas Jefferson Center of Political Economy, University of Virginia, VA.

Vesterlund, Lise. 2006. "Why do People Give?" in Richard Steinberg and Walter W. Powell eds., *The Nonprofit Sector*, 2nd edition, Yale Press.