Multiple Information Signals in the Market for Charitable Donations*

ERICA E. HARRIS, *Rutgers University – Camden*

DANIEL G. NEELY, *University of Wisconsin – Milwaukee*

ABSTRACT

We find evidence indicating that donors use third-party rating information when they donate to U.S. nonprofit organizations (nonprofits). Specifically, using a sample of over 3,800 unique nonprofits rated by the three largest charity rating organizations in 2007, and over 12,000 unrated control nonprofits, we find that rated nonprofits have significantly higher direct donations than unrated charities. We also hypothesize and find that nonprofits with ratings from multiple rating organizations receive incrementally higher levels of donations. In addition, although charities that receive a positive rating have higher levels of donor support than those receiving a negative rating, both positively and negatively rated nonprofits receive a higher level of direct donations than unrated nonprofits. Finally, we find that nonprofits with consistently good ratings receive higher donations than those with mixed or consistently negative ratings, indicating the donor community values consistency across the three rating agencies.

Signaux d'information multiples sur le marché des dons de bienfaisance

RÉSUMÉ

Les observations des auteurs révèlent que les donateurs utilisent les notes de solvabilité attribuées par des tiers dans leurs décisions relatives au versement de dons à des organismes sans but lucratif (OSBL) aux États-Unis. Plus précisément, à partir d'un échantillon d'au-delà de 3 800 OSBL différents, notés par les trois agences de notation des organismes de bienfaisance les plus importantes en 2007, et de plus de 12 000 OSBL non notés servant d'échantillon de contrôle, ils constatent que les OSBL notés reçoivent des dons directs sensiblement plus élevés que les organismes de bienfaisance non notés. Les auteurs posent également l'hypothèse que les OSBL dont les notes proviennent de plusieurs agences de notation reçoivent des dons marginalement plus élevés, hypothèse qui se vérifie. De plus, bien que les organismes de bienfaisance qui reçoivent une note positive affichent des niveaux supérieurs de soutien de la part des donateurs que ceux qui reçoivent une note négative, les OSBL dont les notes sont positives et ceux dont les notes sont négatives reçoivent les uns comme les autres un niveau de dons directs plus élevé que les OSBL qui ne sont pas notés. Enfin, les auteurs constatent que les OSBL qui sont systématiquement bien notés reçoivent des dons supérieurs à ceux dont les notes varient ou sont systématiquement négatives, ce qui indique que la communauté des donateurs attache de la valeur à l'uniformité des notes attribuées par les trois agences de notation.

* Accepted by Vaughan Radcliffe. The authors would like to thank Linda Parsons, Steven Balsam, Michelle Yetman, Andrea Roberts as well as anonymous reviewers and participants at the 2013 American Accounting Association Government and Nonprofit Section meeting and the 2013 American Accounting Association Annual meeting for their thoughtful comments and suggestions for improving this project.

Contemporary Accounting Research Vol. 33 No. 3 (Fall 2016) pp. 989–1012 © CAAA
doi:10.1111/1911-3846.12175

1. Introduction

The nonprofit sector is a large and significant component of the U.S. economy. There were over 1.6 million tax-exempt organizations (nonprofits) registered with the IRS in 2011 (Independent Sector 2012a). The sector contributes $751 billion worth of output, employs roughly 10 percent of the country's workforce, and accounts for 9 percent of the wages paid (Independent Sector 2012b). Charitable giving is an economically prodigious activity with total charitable giving in 2011 estimated to be $298.42 billion (Independent Sector 2012b). The majority of households give to a charitable nonprofit, and the majority of total gifts are given by individuals (rather than corporations or foundations) (Independent Sector 2012b).

While the market for charitable giving is robust, comparatively little information is available to donors about the quality of the nonprofit they donate to. In an attempt to bridge the information gap between prospective donors and nonprofit organizations seeking contributions, several rating agencies have emerged to provide donors with credible signals on the underlying quality of the nonprofit organization. As identified by Lowell, Trelstad, and Meehan (2005), the three prominent U.S. charity rating agencies are Charity Navigator, the American Institute for Philanthropy (AIP), and the Better Business Bureau Wise Giving Alliance (BBB).[1] All three rating agencies provide independent third-party information related to the quality of the nonprofit organizations they review; and while each rating agency approaches the task of generating a quality signal with their own proprietary methodology; all three provide a single rating to aid donors in their donation decision.

Rating agencies primarily rely on publicly available information (i.e., IRS Form 990), therefore their value is in synthesizing available financial data to provide donors with easily accessed and understood information. Prior research has confirmed donors' use of third-party rating information (Sloan 2008; Gordon, Knock, and Neely 2009; Chen 2009; Grant 2010). In particular, earlier studies find that receiving a relatively positive (negative) rating from Charity Navigator is associated with an increase (decrease) in public support (Gordon et al. 2009; Grant 2010), while passing (failing) the BBB twenty standards of accountability is associated with a higher (lower) level of contributions (Sloan 2008; Chen 2009). This study goes beyond these findings by comparing rated nonprofits to unrated nonprofits as a means of examining the information value of charity ratings as a whole, in addition to analyzing rating information from all three national rating agencies simultaneously. This is especially important given that rating agencies are expanding the number of nonprofits they rate, and to date little is known about whether the rating agencies provide complimentary or conflicting signals of charity quality.

Broadly, the purpose of this study is to evaluate the aggregate impact of national rating agencies on the market for charitable contributions. In particular, this study contributes to the literature in three distinct ways. First, we empirically test whether donors give more to nonprofits rated by at least one of the three major rating agencies when compared to unrated nonprofits. Prior studies looking at donor response to rating agencies have focused solely on nonprofits that receive agency ratings. We expand the focus beyond the less than 1 percent of U.S. nonprofit organizations who receive a rating, to consider whether simply being rated leads to an increase in public support.[2] Using multiple rating sources, we are also able to test the incremental impact of receiving multiple ratings.

1. The AIP has rebranded and is now called Charity Watch (AIP 2012). During our sample period, the organization was referred to as the AIP and we will refer to the organization as AIP throughout this article.
2. The fewer than 1 percent represents roughly 8,000 out of 1.6 million tax exempt organizations in the United States.

Second, we consider whether nonprofits receiving a positive (negative) rating receive a higher (lower) level of public support than nonrated nonprofits. While prior literature has found an association between the rating a nonprofit receives and public support, our study is the first to compare receiving a positive or negative rating with receiving no rating at all. Finally, we consider whether the market values signals that are consistent across the three rating agencies; in effect, testing whether a relatively noisy signal is weighed less by donors.

Unfortunately, like many donor contribution studies, our project is limited by the endogenous nature of donor giving. That is, do donors give more based on ratings or are ratings assigned to nonprofits that receive more in contributions? To address the issue of endogeneity in our study we triangulate our results by applying three distinct research methods. First, we control for self-selection using a two-stage Heckman (1979) approach which controls for the likelihood of being rated using a first stage model. Second, we construct a matched-pair sample matching our rated nonprofits with their unrated counterparts based on nonprofit industry, size, and donation reliance. Finally, we employ change analyses to identify sample nonprofits that are newly rated and test the response of new ratings to the change in donations received by the nonprofit.

According to Tucker (2010), in situations where selection bias is caused by unobservable variables (such as specific rating agency selection criteria), a two-stage Heckman inverse Mills ratio (IMR) method is most appropriate for mitigating the effects of selection bias. As a result, we implement a Heckman (1979) two-stage estimation process in our main analyses, and include our matched-pair and change analyses in the robustness section of the paper. The benefit of the Heckman (1979) two-stage process is the ability to control for the likelihood that a nonprofit has been selected using a first-stage prediction model. Our second-stage models, after controlling for the probability of being rated, test four main research questions: donor response to being rated verses unrated, donor response to multiple ratings, donor response to good/bad ratings, and donor response to ratings consistency.

Using an industry-diverse sample of over 16,000 nonprofit organizations with fiscal year 2007 data, we find that nonprofit organizations rated by at least one rating agency receive more donations relative to nonprofits that are unrated. Further we find an incremental increase in donations accruing to nonprofits that receive additional ratings, and that both positive and negative rated nonprofits in fact receive a higher level of direct donations relative to unrated nonprofits. In addition, we document that nonprofits that receive a positive rating are associated with a higher level of public support relative to nonprofits receiving a negative rating. Finally, we find that the market values consistency across the three rating agencies, supporting the notion that noisy signals are valued less by donors.

These findings indicate that the three national rating agencies do influence the market for charitable giving. We believe these results are particularly interesting given that the methodologies used to rate nonprofits vary considerably across the rating agencies. Understanding the credibility of the signals provided by rating agencies as well as the limitations of these signals is important for both donors and regulators alike. Donors aim to make the most informed giving decision, while regulators are focused on enhanced nonprofit transparency. Perhaps, from the perspective of regulators, there should be some concern for the fact that the vast majority of nonprofits are not rated by any third-party rating organization. Absent ratings information from third-party sources, donors are left the difficult task of deciphering Form 990 themselves. In fact, according to research by Hope Consulting, if watchdog groups were effectively able to meet donors information needs "we estimate that $15 billion a year could be shifted to higher-performing nonprofits."[3]

3. Money for Good II research report available at: http://www.hopeconsulting.us/wordpress/wp-content/uploads/2013/01/MFG2-Full_Nov-2011.pdf.

The next section provides background information on the three rating agencies. Section 3 develops and posits our hypotheses; section 4 discusses the sample and methodology. We present study findings in section 5, test of robustness in section 6, and the last section concludes with final remarks.

2. The rating agencies

Charitable nonprofits are accountable to numerous stakeholders including donors, lenders, and community members. Unlike for-profit companies, whose objective function is rather straightforward and whose market is considered efficient, charities are fraught with a complex objective function and a dearth of credible information about their performance (Steinberg 1986; Saxton, Neely, and Guo 2014). Donors base their decision to give on a variety of factors, including affiliation, peer pressure, religious grounds, and sense of common purpose (Gordon and Khumawala 1999). Some donors are rather insensitive to the underlying quality of the nonprofit while other donors are found to allocate their giving based on the perceived quality of the nonprofit organization (Parsons 2007; Li, McDowell, and Hu 2012).

However, unlike for-profit companies with strict reporting requirements and relatively efficient markets processing and pricing the information, historically nonprofits have operated largely free from market oversight. Stakeholders interested in performing an objective assessment of a nonprofit organization have to comb through publicly disclosed informational tax returns (i.e., Form 990) and/or request information directly from the nonprofit organization. The time and relative expertise required to effectively gather and process nonprofit information deters many stakeholders. To fill this informational void in the market for charitable contributions, several third-party organizations adopted accountability standards to measure the effectiveness of charity organizations. The three most prominent organizations are the BBB, Charity Navigator, and the AIP (Lowell et al. 2005).

The BBB Wise Giving Alliance was formed in 2001 as a result of a merger between the National Charities Information Bureau and the Council of Better Business Bureaus' Foundation (BBB 2012). The BBB evaluates just over 1,400 nonprofits against 20 standards for accountability.[4] The BBB chooses to evaluate nonprofits based in part on inquiries from the public as well as requests from national charities (BBB 2014). Each BBB charity report indicates whether a charity meets all 20 standards or discusses why the charity does not meet particular standards (BBB 2012).[5]

While the BBB accredits nonprofit organizations, Charity Navigator and AIP provide a rating of select nonprofits.[6] AIP grades approximately 600 nonprofits along two financial dimensions: percent spent on charitable purpose, and cost to raise $100.[7] AIP chooses to evaluate larger nonprofit organizations (those that receive at least $1 million in public support annually) and are of interest to donors nationally (AIP 2014). Grades range from A (excellent) to F (poor) (AIP 2012). Stakeholders interested in viewing AIP's ratings must first become a member by making a contribution of at least $40 (AIP 2012).

Compared to the relatively limited scope of AIP and BBB, Charity Navigator rates over 5,000 nonprofits. Nonprofits rated by Charity Navigator receive a star rating from zero (the worst) to four stars (the best) based on financial measures derived from the Form 990.[8] Charity Navigator only rates 501(c)(3) nonprofits that receive more than

4. http://www.bbb.org/us/standards-for-charity-accountability.
5. Although for-profit organizations accredited by BBB are required to pay an annual fee, the Wise Giving Alliance of the BBB does not require an annual fee for charity organizations to be included.
6. We refer to the BBB Pass/Fail assessment as a rating throughout the manuscript.
7. http://www.charitywatch.org/criteria.html.
8. www.charitynavigator.org.

$500,000 in public support, have more than $1,000,000 in total revenue, have at least seven years of Form 990s available, report greater than zero fundraising expenses, and are based in the United States (Charity Navigator 2014). Charity Navigator's ratings can be accessed free of charge and were viewed by almost five million donors in 2011 (Charity Navigator 2012). Charity Navigator claims to be the largest and most used third-party charity evaluator (Charity Navigator 2012). The next section motivates and develops our hypotheses.

3. Hypotheses

The information-production hypothesis posits that bond issuers will apply for bond ratings to alleviate investors' concerns about underlying credit quality (Bongaerts, Cremers, and Goetzmann 2012). As stated by Bongaerts et al. (2012, 115), "investors are adverse to uncertainty, which is reduced by adding extra ratings." If the information-production hypothesis holds in the market for charitable contributions, we expect nonprofits rated by one of the three charity evaluators to receive more donations than nonprofits not rated. However, prior literature finds that the majority of donors do not visit a charity evaluator's website before contributing to a nonprofit (Cnaan, Jones, Dickin, and Salomon 2011). Furthermore, donors are motivated to give by factors such as affiliation that are unrelated to whether the nonprofit is rated (Gordon and Khumawala 1999). It is therefore an open empirical question whether rated nonprofits receive more contributions than nonrated nonprofits. As a result, we present our first hypothesis in the null form:

HYPOTHESIS 1: *Rated nonprofits receive no more or less donations than their nonrated peers.*

Also consistent with the information-production hypothesis, we expect that the more sources of rating coverage a nonprofit receives, the more donations the nonprofit will accumulate. This builds on our first hypothesis and follows bond-rating literature which confirms a certification theory for why companies go after multiple bond ratings (Bongaerts et al. 2012). In particular, Bongaerts et al. (2012) find that multiple credit ratings play an important role in companies that have split ratings between junk and investment grade, concluding that additional ratings help to certify their creditworthiness in the bond market. We also draw support for our second hypothesis from the security analyst literature. In particular, Brennan, Jegadeesh, and Swaminathan (1993) report that stocks with greater analyst coverage react faster to market-wide information compared to those with less analyst coverage, indicating that multiple analysts' forecasts aid stockholders in their investment decisions. In addition, prior theoretical research finds that multiple signals exhibit complementary precision compared with being individually disclosed (Lundholm 1991). However, identical to our first hypothesis, if donors do not value rating information at all, we would not expect an incremental benefit from being rated by more than one charity evaluator. Accordingly, we state our second hypothesis in the null form:

HYPOTHESIS 2: *Additional ratings are not associated with donations.*

Gordon et al. (2009) and Grant (2010) find that higher Charity Navigator ratings are associated with more direct donations. Likewise, Chen (2009), and Sloan (2008) find that nonprofits that pass all 20 of the BBB standards accrue more public support relative to those nonprofits that fail at least one BBB standard. However, our study is the first to consider how a group of charities not rated by any of the three large third-party rating agencies compares with nonprofits receiving relatively good or poor ratings. The importance of this distinction cannot be overstated. Prior capital markets literature finds that companies voluntarily disclose bad news and that withholding information has negative consequences (Milgrom 1981; Skinner 1994). Similarly, evidence in the debt market indicates that when companies go public, those without credit ratings are underpriced significantly more than companies with established credit ratings (An and Chan 2008). This research concludes that the existence of a credit rating reduces investors' uncertainty about company value by providing useful information believed to reduce information asymmetries in the IPO market. If these findings hold in the nonprofit sector, we would expect even nonprofits that fail the BBB standards, receive a low Charity Navigator rating and/or receive a low AIP grade to receive more in public support than nonprofits not assessed by any of the third-party accountability organizations.

Consequently, our third hypothesis is twofold. First, we expect nonprofits with both good and bad ratings to attract more donations over unrated nonprofits. Second, consistent with prior literature, we expect that nonprofits with good ratings will attract more public support than nonprofits with bad ratings. Hence, our third hypothesis is stated as follows:

Hypothesis 3a: *Good or bad rating signals are associated with more donations than no rating signal.*

Hypothesis 3b: *Good rating signals are associated with more donations than bad rating signals.*

Finally, a condition of a signal's credibility is the degree to which multiple signals are consistent (Chung and Kalnins 2001; Fischer and Reuber 2007; Connelly, Certo, Ireland, and Reutzel 2011). Split bond ratings have been found to increase the borrowing costs of issuers (Livingston and Zhou 2010). In addition, prior research has shown that investors purchasing stocks with the most favorable consensus analyst recommendations can make positive abnormal returns (Barber, Lehavy, McNichols, and Trueman 2001). We posit that if nonprofits evaluated by more than one charity rating organization receive a consistent signal about their underlying quality (either good or bad) donors will respond by relying more on the signal and thus strengthen the association between public support and the rating. However, if the signal is considered noisy (i.e., inconsistent) we would expect donors to weigh the signal less and give less (Verrecchia 2001). Therefore, our fourth hypothesis is stated as follows:

Hypothesis 4: *Rating signals that are consistent across rating agencies exhibit a stronger association with donations.*

The next section discusses our sample and research methodology.

4. Research methodology and sample

As mentioned earlier, in our setting, a concern is whether the process rating agencies use to determine which nonprofits they rate biases for finding a relation between being rated and contributions. That is, if rating agencies only rate nonprofits that receive more in public contributions, we would expect a positive association between being rated and contributions. One common solution for mitigating self-selection bias is the use of instrumental variables; however, we are unable to identify a suitable variable which meets the criteria necessary to conduct that type of analyses in our study sample (Larcker and Rusticus 2010). Therefore, to address the issue of self-selection in our sample, we employ a Heckman (1979) two-stage estimation process. Tucker (2010) confirms the ability of a two-stage Heckman IMR method to properly mitigate the effect of selection bias in samples where unobservable variables (such as specific rating agency selection criteria) are present. The IMR method is also particularly suited for binary treatment choices, such as whether a nonprofit is rated or not, where the first stage model is estimated as a choice model. Specifically, our first stage model includes covariates believed to predict which nonprofit organizations are selected for review by the three national rating agencies. We include size and industry, in addition to two thresholds set by the rating agencies: total contributions greater than \$500,000 as well as total revenues more than \$1 million. Charity Navigator requires that nonprofits have at least \$500,000 in total contributions and \$1 million in total revenues to be considered for evaluation, while AIP rates nonprofits that have at least \$1 million in public support.[9] In sum, our first stage model to predict being rated by AIP, BBB, or Charity Navigator (with entity subscripts suppressed) is:

$$\begin{aligned} Rated_t = \alpha + \beta_1 Total\ Assets_t + \beta_2 TC > \$500k_t + \beta_3 TR > \$1M_t \\ + industry\ fixed\ effects + \varepsilon_t, \end{aligned} \tag{1}$$

where $TC > \$500k$ and $TR > \$1M$ are indicator variables equal to one for nonprofits meeting the total contributions and total revenues requirements, respectively, and all other model variables are defined below.[10]

We use equation (1) to calculate the inverse Mills ratio (IMR) as the ratio of the standard normal probability density function (p.d.f.) over standard normal cumulative density function (c.d.f.). The resulting IMR variable is then included in the second-stage regression as a bias correction term in addition to other model control variables. That is, our second-stage model tests the impact of ratings on nonprofit contributions, after controlling for the probability of being rated.

In particular, we include IMR, calculated from equation (1), in the standard donations demand model used extensively in both the accounting and economics literature (Weisbrod and Dominguez 1986) to estimate the following model (with entity subscripts suppressed):

$$\begin{aligned} Direct\ Donations_t = \alpha + \beta_1 Ratings\ Test\ Variable_{t-1} + \beta_2 IMR_t + \beta_3 PRICE_{t-1} \\ + \beta_4 Fundraising\ Expenses_{t-1} + \beta_5 Age_t + \beta_6 Total\ Assets_{t-1} \\ + \beta_7 Program\ Service\ Revenues_{t-1} + \beta_8 Other\ Revenue_{t-1} \\ + \beta_9 Government\ Grants_{t-1} + industry\ fixed\ effects + \varepsilon_t, \end{aligned} \tag{2}$$

where our model is estimated using robust regression techniques (iteratively reweighted least squares), which assigns a weight to each observation with higher weights given to observations which meet the assumptions underlying standard multiple regression.[11] Robust regressions also adjust for data outliers identified as a potential problem when working with IRS Form 990 data, from which we draw our sample (Tinkelman and Neely 2011).

9. In our main model we include an indicator variable for nonprofits with over \$500 thousand in total contributions following the threshold set by Charity Navigator. However, when we replace this variable with an indicator for nonprofits with \$1 million in direct donations (public support) following the threshold set by AIP our results are unchanged (i.e., our test variable *Rated* is still significantly different from zero with a coefficient of 0.649). Because 79 percent of our sample is rated by Charity Navigator, versus 7 percent rated by AIP, we use the Charity Navigator threshold in tabulated results. Our results are also robust to restricting our sample to nonprofits with direct donations (rather than total contributions) greater than \$500,000 and total revenues more than \$1 million (i.e., our test variable *Rated* is still significantly different from zero with a coefficient of 0.751).

10. In untabulated analyses we replace our total contributions and total revenues indicator variables with logged continuous variables separately and together and find consistent results.

11. UCLA Institute for Digital Research and Education. (http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter4/statareg4.htm).

Model variables

Our response variable, *Direct Donations* or direct public support, has been defined by the IRS as: "amounts of contributions, gifts, grants, and bequests that the nonprofit received directly from the public." We focus on this measure of public support given that we aim to understand the impact charity ratings have on individual giving as opposed to indirect sources such as the United Way or government grantors. Specifically we measure *Direct Donations* from the 2008 revised Form 990 as total contributions (line 1h) less federated campaigns (line 1a) and government grants (line 1e).[12]

Following prior literature (e.g., Trussel and Parsons 2007) we specify our response variable at time *t*, while our test and control variables are lagged (with the exception of *Age*). We believe this is the most appropriate form of our model given that ratings may be issued at any time during the fiscal year and that donors need time to investigate and incorporate rating information into their donation decisions.

In our primary analysis (Hypothesis 1) we define our ratings test variable, *Rated*, as an indicator variable equal to one for nonprofit organizations rated by Charity Navigator, BBB, or AIP.[13] All ratings are measured based on the published date of 2007. The rating agencies typically base their ratings on financial information from the prior year (2006 for our study). As discussed in more detail below, some overlap exists between the three rating agencies, which allows us to extend our analysis to address the effect of additional ratings on donor contributions. In particular, Hypothesis 2 test variables, *Two Ratings* and *Three Ratings*, take on the value of one for nonprofits rated by two or all three of the rating agencies, respectively. Note that in our Hypothesis 2 model we restrict our sample to nonprofits that are rated, given that our aim is to assess whether additional ratings (*Two Ratings* or *Three Ratings*) provide an incremental value above and beyond simply being rated.

For our third hypothesis, the impact of positive or negative ratings, we incorporate the actual ratings assigned to treatment nonprofits in our sample. Here we define the test variable *Good* equal to one for nonprofits who receive a rating in the top half of the rating definitions from each agency. That is, three or four stars (out of four stars) from Charity Navigator, a grade of B- or better from AIP, or a pass rating from BBB, and zero otherwise. We define the test variable *Bad* equal to one for nonprofits that receive zero or one star from Charity Navigator, a grade of C or worse from AIP, or a fail rating from BBB, and zero otherwise.[14] In the case of nonprofits that receive multiple ratings from the same agency in the same year (in the case of BBB and AIP), we use the latest 2007 rating assigned to the charity.

Finally, to test the consistency of ratings across the three rating agencies (Hypothesis 4) we define our final set of test variables: *Consistent_good, Consistent_bad,* and *Mixed*. *Consistent_good* is an indicator variable activated for nonprofits receiving multiple ratings all classified as *Good,* while our *Consistent_bad* test variable is equal to one for nonprofits with multiple ratings classified as all *Bad* ratings.[15] A nonprofit that has at least two ratings and has received a combination of both *Good* and *Bad* ratings is coded as our *Mixed*

12. We note that our results are unchanged when we define our response variable, *Direct Donations*, as 2008 Form 990 lines 1c (special events) plus 1f (all other contributions).
13. In 2007 BBB published four separate quarterly magazines. We code an agency as *Rated* if they appear in any one of these publications and have been assigned a pass or fail rating. That is, we do not consider a BBB nonprofit *Rated* if it is classified as "under review" or "unable to verify." AIP publishes ratings twice a year and nonprofits rated in either time period are coded as *Rated*.
14. We code 46 observations equal to zero for nonprofits that receive both a *Good* and *Bad* rating.
15. Our consistency variables differ from the *Good*/*Bad* test variables in that they exclude nonprofits that receive only one rating. That is, for a nonprofit to take on a value as either a *Consistent_good* or *Consistent_bad* nonprofit, it must receive two or three ratings.

variable equal to one, and zero otherwise. In addition, to isolate consistent ratings received from multiple rating agencies, beyond a single rating, we also define *Single Rating* equal to one for nonprofits that receive a single rating, and zero otherwise. As a result, our consistency variables are able to disentangle the effect of receiving ratings from multiple rating agencies from single *Good*/*Bad* ratings received from one particular rating source.

As an example of a nonprofit that has received both a *Good* and *Bad* rating, we have reproduced the ratings for a sample nonprofit, Relief International, in the Appendix. Relief International received a four star rating from Charity Navigator for 2007. However, Relief International failed to meet all 20 BBB standards of accountability. Specifically, while Relief International's reported financial information from the Form 990 supported the high rating from Charity Navigator, Relief International failed two BBB standards: standard 13 which requires accurate expense reporting, and standard 14 which requires a board-approved budget detailing functional expenses. In addition, BBB could not verify that the nonprofit spent no more than 35 percent of related contributions on fundraising (standard 9). This example highlights the fact that third-party evaluators apply differential weights to transparency, accuracy, and reported results culminating at times in contrasting ratings.

Turning to our control variables, prior nonprofit literature has studied the determinants of contributions and found several variables to be significant predictors. Specifically, Trussel and Parsons (2007) describe efficiency, information quantity, and information quality as constructs found to be significantly associated with nonprofit donor contributions. In addition, several studies have included the notion of donations being "crowded out" by other sources of revenue (Weisbrod and Dominguez 1986; Posnett and Sandler 1989; Callen 1994; Okten and Weisbrod 2000).

To control for the effect of nonprofit efficiency on nonprofit contributions, we include lagged *PRICE* following the findings of Weisbrod and Dominguez (1986) as well as subsequent studies which confirm this relationship (Posnett and Sandler 1989; Greenlee and Brown 1999; Baber, Roberts, and Visvanathan 2001; Parsons 2003; Tinkelman 2004; Trussel and Parsons 2007; Tinkelman and Mankaney 2007). *PRICE* is operationalized as the inverse of the program ratio, or 1/(Program Expense/Total Expense).

Information quantity represents the amount of information displayed to potential contributors, typically in the form of advertising or other means of making the mission of a nonprofit public. We include lagged *Fundraising Expenses* to proxy for the amount of information made available to contributors following prior literature (Weisbrod and Dominguez 1986; Tinkelman 1999).

Information quality is considered paramount in the decision to contribute to a nonprofit. Trussel and Parsons (2007) identify two variables that are found to be related to the information quality of a nonprofit in the decision to donate: age and size. *Age* is operationalized as the number of years since the nonprofit's initial 501(c)(3) filing for tax exempt status. Lagged *Total Assets* (IRS Form 990, line 59) is included to proxy for nonprofit size.[16]

Several papers have also documented the notion of donors refraining from making donations to nonprofits who receive high levels of government grants, program service revenues, or other revenues, which are considered to "crowd out" donations (Weisbrod and Dominguez 1986; Posnett and Sandler 1989; Callen 1994; Okten and Weisbrod 2000). However, other literature has documented the "crowding-in" effect of these same variables (Okten and Weisbrod 2000; Petrovits, Shakespeare, and Shih 2011). As such, we include

16. Following the guidance of Lennox, Francis, and Wang (2012), we confirm the robustness of our size measure finding comparable first and second stage results when we define size in terms of total contributions or total revenues.

these three lagged variables to control for the presence of income sources considered either to be substitutes or complements by the donor population. However, given the mixed results of prior literature, we do not make sign predictions for these variables. A description of our model variables is included in Table 1.

Sample

We construct our sample of nonprofit organizations from the 2007 listings of rated nonprofits by the three national rating agencies: Charity Navigator, BBB and AIP.[17] Between the three agencies we are able to identify complete rating information for 3,861 unique nonprofit organizations. Charity Navigator rated the largest number of nonprofit organizations in our sample with 3,555 nonprofit ratings, while BBB and AIP provided ratings

TABLE 1
Variable definitions

Direct Donations	Log of 2008 Form 990 line 1h (Total Contributions) less line 1a (Federated Campaigns) and line 1e (Government Grants)
Rated	1 for nonprofits rated by either BBB, AIP, or Charity Navigator; 0 otherwise
Two Ratings	1 for nonprofits receiving two ratings; 0 otherwise
Three Ratings	1 for nonprofits receiving three ratings; 0 otherwise
Good	1 for an nonprofits that receive 3 or 4 stars from Charity Navigator, a grade of B- or better from AIP, or a passing grade from BBB; 0 otherwise
Bad	1 for nonprofits that receive 0 or 1 star from Charity Navigator, a grade of C or worse from AIP, or a failing grade from BBB; 0 otherwise
Single Rating	1 for nonprofits that receive a single rating; 0 otherwise
Consistent_good	1 for nonprofits that receive two or more ratings and all ratings are *Good*; 0 otherwise
Consistent_bad	1 for nonprofits that receive two or more ratings and all ratings are *Bad*; 0 otherwise
Mixed	1 for nonprofits that receive two or more ratings and a combination of *Good* and *Bad* ratings; 0 otherwise
IMR	Inverse Mills ratio calculated from equation (1)
PRICE	1/(Program Service Expenses from line 14/Total Expenses from line 17)
Fundraising Expenses	Log of Form 990 line 15
Age	Number of years since the nonprofit's IRS tax-exemption ruling
Total Assets	Log of Form 990 line 59
Program Service Revenues	Log of Form 990 line 2
Other Revenue	Log of Form 990 line 11
Government Grants	Log of Form 990 line 1c
TC > $500k	1 for nonprofits with total contributions greater than $500,000; 0 otherwise
TR > $1M	1 for nonprofits with total revenues greater than $1,000,000; 0 otherwise

Notes:

Unless otherwise noted, Form 990 references are to the old form (pre-2008), as our sample is drawn from 2007.

17. We select the year 2007 for our analysis following prior literatures (e.g., Trussel and Parsons 2007) that suggest that donors need at least one year to incorporate information into their giving decision. Based on these literatures and data availability at the time of data collection, 2007 data are the most up to date information obtainable for our study.

TABLE 2
Sample

Panel A: Distribution of ratings across rating agencies

	AIP	BBB	Charity Navigator	Total
1 rating	48	193	3,141	3,382
2 ratings: AIP & BBB	65	65	–	130
2 ratings: AIP & CN	87	–	87	174
2 ratings: BBB & CN	–	182	182	364
3 ratings: AIP, BBB, & CN	145	145	145	435
Total	345	585	3,555	4,485
Less overlap in ratings (65 + 87 + 182 + 145 + 145)				−624
Total unique rated nonprofits				3,861
Control sample of unrated nonprofits				12,846
Total sample				16,707

Panel B: Distribution of ratings across nonprofit industries

Nonprofit industry	Rated	Percent	Unrated	Percent	Total	Percent
Health	494	12.8	3,900	30.4	4,394	26.3
Human services	898	23.3	3,454	26.8	4,352	26.0
Other industries	1,662	43.0	2,102	16.4	3,764	22.5
Education	287	7.4	2,647	20.6	2,934	17.6
Arts, culture, and humanities	520	13.5	743	5.8	1,263	7.6
Total	3,861	100.0	12,846	100.0	16,707	100.0

Notes:

Other industries include: environment, public and societal benefit, religion, international, mutual benefit, and unknown.

for 585 and 345 charities, respectively.[18] Given that these rating agencies work independently, overlap in ratings is present in our sample. Sixty-five nonprofits were rated by both BBB and AIP, 182 by BBB and Charity Navigator, and 87 by AIP and Charity Navigator. In addition, 145 nonprofit organizations in our sample were rated by all three rating agencies. This leaves 48 nonprofits that were uniquely rated by AIP, 193 reviewed only by BBB, and 3,141 nonprofits solely rated by Charity Navigator.[19] Table 2, panel A tabulates these figures and provides a reconciliation of the full sample.

To allow our analysis to relate the impact of rating information to direct donations, we merge rating records with financial data drawn from IRS Form 990 made available in digital format from the National Center for Charitable Statistics (NCCS). Specifically we

18. The reason for the difference between total nonprofits rated by the rating agencies today (approximately 8,000) and total rated nonprofits in our sample (4,485) is twofold. First, we look at 2007 ratings and the agencies (especially Charity Navigator) have increased the number of nonprofits they rate over the years as funding for their work increases. Second, we are only able to include nonprofits for which we have all model variables. While we do our best to fill in digitized data with hand-collected financial information, however, for nonprofits that file a Form 990 EZ, necessary financial information is not readily available for collection.

19. Given the large percentage of our sample rated by Charity Navigator, in untabulated analyses we confirm that our first and second stage model results are robust when we define our sample in terms of each rating agency separately, when we exclude nonprofits rated by BBB and AIP, and when we alternatively group just BBB and AIP nonprofits.

TABLE 3
Descriptive statistics

Sample		Rated	Unrated	AIP	BBB	Charity Navigator	Full sample
N		3,861	12,846	345	585	3,555	16,707
Direct Donations$_t$	Mean	12,500	5,214	49,000	31,200	10,700	6,900
	SD	52,100	38,300	150,000	93,100	35,100	42,000
PRICE$_{t-1}$	Mean	1.371	1.593	1.497	1.834	1.283	1.542
	SD	5.012	10.465	1.363	12.829	0.570	9.481
Fundraising Expenses$_{t-1}$	Mean	1,097	424	4,479	2,873	946	579
	SD	4,112	2,585	11,300	9,061	3,312	3,021
Age$_t$	Mean	32.687	31.498	31.252	25.506	33.529	31.772
	SD	19.010	21.619	17.859	17.877	18.909	21.051
Total Assets$_{t-1}$	Mean	52,400	146,000	72,500	83,200	52,300	125,000
	SD	341,000	859,000	326,000	544,000	348,000	772,000
Program Service Revenues$_{t-1}$	Mean	4,960	56,200	9,930	8,459	4,890	44,300
	SD	54,900	339,000	124,000	98,700	56,400	299,000
Other Revenue$_{t-1}$	Mean	285	1,195	406	655	264	984
	SD	3,232	7,962	1,359	5,190	3,040	7,163
Government Grants$_{t-1}$	Mean	5,104	4,032	13,300	14,000	4,276	4,112
	SD	29,400	48,400	35,200	41,900	26,800	47,200

Notes:

Reported values are in $thousands, except for *PRICE*, which is measured in dollars, and *Age*, which is measured in years. Raw values are presented for illustrative purposes only; in our multivariate models, we use the log form of these variables—as defined in Table 1.

use the NCCS Statistics of Income (SOI) file to draw financial data for our rated agencies, as well as provide unrated control nonprofits. The SOI file provides for 12,846 control nonprofits, for a total sample size of 16,707 unique nonprofits.[20] While the SOI file does not include every nonprofit for every year (as reported earlier, the sector included 1.6 million registered tax-exempt entities in 2011), it does reflect "over 90 percent of all nonprofit revenues," according to Yetman and Yetman (2013, 1049).

Table 2, panel B presents the distribution of the nonprofit industries in our sample. We find that the largest percentage (26.3 percent) of our sample is made up of Health nonprofits; this is followed by human services (26.0 percent), other industries (22.5 percent), education (17.6 percent), and arts (7.6 percent). Panel B also reports industry classifications for our rated and unrated subsamples. Here we find that other industries make-up the largest percentage of rated nonprofits (43.0 percent), followed by human services (23.3 percent), arts (13.5 percent), health (12.8 percent), and education (7.4 percent). While the largest percentage of unrated nonprofits (30.4 percent) are classified as health nonprofits, followed by human services (26.8 percent), education (20.6 percent), other industries (16.4 percent), and arts (5.8 percent).[21]

20. The SOI file includes data for nonprofits with assets of $50 million or more as well as a random sample of smaller nonprofits (Feng, Ling, Neely, and Roberts 2014). In addition, nonprofits with less than $50,000 in gross receipts are not required to file a Form 990.

21. Given that our rated and unrated samples are unbalanced in terms of hospitals and universities, in our sensitivity analyses we also run our models excluding these two industry classifications.

TABLE 4
Correlation table

	log Direct Donations$_t$	*Rated$_{t-1}$*	*PRICE$_{t-1}$*	*log Fundraising Expenses$_{t-1}$*	*log Age$_t$*	*log Total Assets$_{t-1}$*	*log Program Service Revenues$_{t-1}$*	*log Other Revenue$_{t-1}$*	*log Gov Grants$_{t-1}$*
log Direct Donations$_t$	1.000	0.435	−0.437	0.075	0.705	0.253	0.296	0.017	0.204
		0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.029*	*0.000*
Rated$_{t-1}$	0.346	1.000	−0.599	0.206	0.473	0.046	−0.205	−0.245	−0.042
	0.000		*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
PRICE$_{t-1}$	−0.298	−0.599	1.000	−0.149	−0.393	−0.016	0.174	0.264	0.054
	0.000	*0.000*		*0.000*	*0.000*	*0.043*	*0.000*	*0.000*	*0.000*
log Fundraising Expenses$_{t-1}$	−0.029	−0.010	0.014	1.000	0.192	0.047	−0.128	−0.134	0.007
	0.000	*0.203*	*0.070*		*0.000*	*0.000*	*0.000*	*0.000*	*0.402*
log Age$_t$	0.591	0.480	−0.410	−0.017	1.000	0.282	0.188	0.014	0.155
	0.000	*0.000*	*0.000*	*0.033*		*0.000*	*0.000*	*0.075*	*0.000*
log Total Assets$_{t-1}$	0.235	0.078	−0.034	−0.021	0.246	1.000	0.333	0.332	0.286
	0.000	*0.000*	*0.000*	*0.006*	*0.000*		*0.000*	*0.000*	*0.000*
log Program Service Revenues$_{t-1}$	0.163	−0.149	0.124	−0.021	0.121	0.318	1.000	0.564	0.447
	0.000	*0.000*	*0.000*	*0.007*	*0.000*	*0.000*		*0.000*	*0.000*
log Other Revenue$_{t-1}$	0.006	−0.175	0.197	−0.029	−0.012	0.269	0.414	1.000	0.450
	0.425	*0.000*	*0.000*	*0.000*	*0.111*	*0.000*	*0.000*		*0.000*
log Gov Grants$_{t-1}$	0.172	−0.018	0.037	−0.017	0.114	0.258	0.403	0.361	1.000
	0.000	*0.019*	*0.000*	*0.032*	*0.000*	*0.000*	*0.000*	*0.000*	

Notes:

Pearson correlations appear above the diagonal, and Spearman below. Italicized *p*-values appear below the correlations. See Table 1 for variable definitions.

TABLE 5
Probit results—first stage model

Dependent variable: $Rated_t$	Coefficient p-value
Constant	1.523
	0.000
log Total Assets$_t$	–0.233
	0.000
$TC > \$500k_t$	1.875
	0.000
$TR > \$1M_t$	0.699
	0.000
Industry fixed effects	Yes
N	16,707
Pseudo adjusted R^2	0.3272
Model Chi-squared	5,910.70
	0.000
Highest VIF	1.89

Notes:

See Table 1 for variable definitions.

In untabulated analysis, we find that industry distributions are approximately the same for each of the three rating organizations individually, with the exception of two industries: arts and international nonprofits. Charity Navigator rates a much larger percentage of arts, culture, and humanities nonprofits (14 percent) than does BBB (2 percent) and AIP (1 percent). Conversely, 22 percent of AIP and BBB rated nonprofits are classified as international nonprofits while Charity Navigator does not rate these type of nonprofits.

Table 3 provides descriptive statistics for model variables in six different compositions of our sample: rated, unrated, AIP, BBB, Charity Navigator, and our full sample. Here we find that mean direct donations are highest for the subset of charities rated by AIP, followed by BBB, and then Charity Navigator. Mean direct donations for the overall rated group are \$12.5 million versus just over \$5 million for the unrated group. This difference is statistically significant at conventional levels ($p < 0.01$) indicating that nonprofits rated by the three major U.S. rating agencies, receive significantly more direct donations than their unrated counterparts. However, we note that the unrated charities in our sample appear to be significantly larger ($p < 0.01$) in terms of total assets with means of \$146 million versus \$52 million for the unrated and rated subsamples respectively.[22]

Table 4 presents Pearson and Spearman correlations for our model variables. Here we find that our test variable, *Rated*, is positively correlated with our response variable *Direct Donations* in support of our main hypothesis. Consistent with prior literature (Trussel and Parsons 2007; Weisbrod and Dominguez 1986) we also find that size and age are significantly correlated with the remaining variables in our model. As suggested by the work of Lennox et al. (2012) we report the highest variance inflation factor (VIF) for each of the first and second stage models presented in Tables 5–9. Consistent with statistical guidance which indi-

22. Note that unrated charities are more likely to include hospitals and educational institutions than rated ones (see industry distributions in Table 2); in our additional analyses presented in Table 7 we confirm our results when we exclude these industries.

TABLE 6
Regression results—tests of hypotheses

Dependent variable: *log Direct Donations$_t$*	(1) Hypothesis 1 Coefficient *p*-value	(2) Hypothesis 2 Coefficient *p*-value	(3) Hypothesis 3 Coefficient *p*-value	(4) Hypothesis 4 Coefficient *p*-value
Constant	6.458	6.404	6.608	6.393
	0.000	0.000	0.000	0.000
Rated$_{t-1}$	**0.684**		**0.452**	
	0.000		***0.000***	
Two Ratings$_{t-1}$		**0.195**		
		0.000		
Three Ratings$_{t-1}$		**0.341**		
		0.000		
Good$_{t-1}$			**0.365**	
			0.000	
Bad$_{t-1}$			**0.119**	
			0.085	
Single Rating$_{t-1}$				**0.755**
				0.000
Consistent_good$_{t-1}$				**1.659**
				0.000
Consistent_bad$_{t-1}$				**1.234**
				0.000
Mixed$_{t-1}$				**1.093**
				0.000
IMR$_t$	−0.542		−0.551	−0.880
	0.000		0.000	0.000
PRICE$_{t-1}$	−0.012	−1.109	−0.012	−0.012
	0.000	0.000	0.000	0.000
log Fundraising Expenses$_{t-1}$	0.117	0.573	0.117	0.095
	0.000	0.000	0.000	0.000
log Age$_t$	−0.091	−0.183	−0.086	−0.079
	0.000	0.000	0.000	0.000
log Total Assets$_{t-1}$	0.462	0.203	0.453	0.494
	0.000	0.000	0.000	0.000
log Program Service Revenues$_{t-1}$	−0.046	−0.018	−0.045	−0.039
	0.000	0.000	0.000	0.000
log Other Revenue$_{t-1}$	0.009	0.000	0.010	0.008
	0.000	0.980	0.000	0.000
log Government Grants$_{t-1}$	0.020	−0.010	0.019	0.011
	0.000	0.000	0.000	0.000
Industry fixed effects	Yes	Yes	Yes	Yes
N	16,707	3,860	16,707	16,707
Adjusted R^2	0.5584	0.7870	0.5595	0.5985
Model *F*-statistic	1,883.16	1,097.86	1,639.96	1,557.51
	0.000	0.000	0.000	0.000
Highest VIF	1.95	1.84	3.95	1.95

Notes:

See Table 1 for variable definitions. Bold represents variables of interest.

cates that VIFs less than 10 are at an acceptable level (Marquardt 1980; Gujarati 1995), all model VIFs are appropriate with a highest reported value of 3.95 (Table 6, column 3).

5. Empirical results

Probit results for our first stage model, which predicts the probability of being rated by one of the three national rating agencies, are presented in Table 5. As expected, we find that size (*Total Assets*), as well as indicator variables for total contributions greater than \$500,000 and total revenues greater than \$1 million, are significantly related to the probability of being rated. It is notable that the coefficient on our size variable (*Total Assets*) is negative suggesting that relatively smaller nonprofit organizations are more likely to receive a third-party rating.

Second-stage model results for Hypothesis 1 are offered in column (1) of Table 6. Consistent with our expectation, nonprofits rated by Charity Navigator, BBB, or AIP receive significantly more direct donations than unrated charities after controlling for the likelihood of being rated. Column 2 presents our test of Hypothesis 2, the impact of multiple ratings on nonprofit contributions. Here we isolate the subsample of rated nonprofits and document the incremental impact of receiving two or three ratings above and beyond being rated. That is, nonprofits receiving two and three ratings receive additional donations over those classified as being rated. In addition, we find a significant difference between our *Two Ratings* and *Three Ratings* test variables ($p = 0.0479$) indicating that significantly more donations accrue to nonprofits rated by all three rating agencies.

Column (3) tests Hypothesis 3, which predicts the differential response of donors to specific information provided by rating agencies (i.e., good or bad ratings). Consistent with our prediction, we find that charities that receive both good and bad ratings receive significantly more direct donations than nonprofits that are unrated. However, the coefficient on our *Good* test variable (0.365) is approximately three times more than our *Bad* variable (0.119). Using a Wald test, we confirm that the coefficient on *Good* is significantly different from *Bad* ($p < 0.01$), indicating that nonprofits that receive a positive rating do in fact receive significantly more direct donations than both unrated and poorly rated charities.

Our final analysis is presented in column (4) of Table 6. The results of this model indicate that ratings consistency is also important to donors. The coefficient on our *Consistent_good* test variable is positive and significantly different from zero demonstrating that the nonprofits who receive *Consistent_good* ratings, beyond being simply rated, accumulate significantly more direct donations. We also find that nonprofits receiving a single rating, those with consistently bad ratings, as well as those with mixed ratings, receive more contributions than their unrated counterparts. However, the *Consistent_good* model coefficient is significantly greater than each of the three remaining test variables ($p < 0.05$). These results are consistent with our predictions and specifically confirm Hypothesis 4 which posits that consistent ratings across rating agencies exhibit a stronger association with public support.

Turning to our control variables, we find consistent significant positive relations between our response variable, *Direct Donations*, and our *Fundraising Expenses, Total Assets, Other Revenue,* and *Government Grants* control variables. These relations are consistent with prior literature. We find a negative sign on our *IMR* control variable which indicates that the error terms in our first and second stage models are negatively correlated. That is, the unobserved factors that make being rated more likely tend to be associated with lower direct donations.

We also find negative relations with *PRICE*, *Age*, and *Program Service Revenues*. While the former is also consistent with prior literature and the latter is consistent with crowding-out, the negative coefficient on *Age* is surprising. However, from Table 4, *Age* and *Total Assets* are highly correlated (significant correlation with coefficient of 0.282). When we exclude *Total Assets* from our model, *Age* becomes positive and significantly different from zero, consistent with prior literature.

TABLE 7
Robustness—sample excluding hospitals and universities

Dependent variable: *log Direct Donations$_t$*	Hypothesis 1 Coefficient *p*-value	Hypothesis 2 Coefficient *p*-value	Hypothesis 3 Coefficient *p*-value	Hypothesis 4 Coefficient *p*-value
Constant	7.322	6.548	7.665	7.487
	0.000	0.000	0.000	0.000
Rated$_{t\text{-}1}$	**0.632**		**0.296**	
	0.000		***0.000***	
Two Ratings$_{t\text{-}1}$		**0.104**		
		0.046		
Three Ratings$_{t\text{-}1}$		**0.334**		
		0.000		
Good$_{t\text{-}1}$			**0.541**	
			0.000	
Bad$_{t\text{-}1}$			**0.247**	
			0.001	
Single Rating$_{t\text{-}1}$				**0.697**
				0.000
Consistent_good$_{t\text{-}1}$				**1.600**
				0.000
Consistent_bad$_{t\text{-}1}$				**1.091**
				0.000
Mixed$_{t\text{-}1}$				**0.988**
				0.000
IMR$_t$	−0.519	−0.254	−0.543	−0.898
	0.000	0.000	0.000	0.000
PRICE$_{t\text{-}1}$	−0.011	−1.258	−0.011	−0.011
	0.000	0.000	0.000	0.000
log Fundraising Expenses$_{t\text{-}1}$	0.134	0.562	0.133	0.110
	0.000	0.000	0.000	0.000
log Age$_t$	−0.136	−0.180	−0.127	−0.123
	0.000	0.000	0.000	0.000
log Total Assets$_{t\text{-}1}$	0.399	0.237	0.378	0.424
	0.000	0.000	0.000	0.000
log Program Service Revenues$_{t\text{-}1}$	−0.037	−0.017	−0.035	−0.033
	0.000	0.000	0.000	0.000
log Other Revenue$_{t\text{-}1}$	0.015	0.000	0.014	0.011
	0.000	0.967	0.000	0.000
log Government Grants$_{t\text{-}1}$	0.019	−0.012	0.017	0.008
	0.000	0.000	0.000	0.005
Industry fixed effects	Yes	Yes	Yes	Yes
N	9,379	3,079	9,379	9,379
Adjusted R^2	0.5932	0.7968	0.5973	0.6344
Model *F*-statistic	1,453.15	1,006.51	1,254.06	1,163.92
	0.000	0.000	0.000	0.000
Highest VIF	2.40	2.48	3.38	2.41

Notes:

See Table 1 for variable definitions. Bold represents variables of interest.

TABLE 8
Robustness—matched sample

Dependent variable: *log Direct Donations*$_t$	Hypothesis 1 Coefficient *p*-value	Hypothesis 3 Coefficient *p*-value	Hypothesis 4 Coefficient *p*-value
Constant	6.029	6.179	6.253
	0.000	0.000	0.000
Rated$_{t\text{-}1}$	**0.754**	**0.625**	
	0.000	***0.000***	
Good$_{t\text{-}1}$		**0.434**	
		0.000	
Bad$_{t\text{-}1}$		**0.386**	
		0.002	
Single Rating$_{t\text{-}1}$			**0.698**
			0.000
Consistent_good$_{t\text{-}1}$			**1.448**
			0.000
Consistent_bad$_{t\text{-}1}$			**0.558**
			0.000
Mixed$_{t\text{-}1}$			**0.949**
			0.000
IMR$_t$	−0.375	−0.390	−0.395
	0.000	0.000	0.000
PRICE$_{t\text{-}1}$	−0.014	−0.014	−0.014
	0.000	0.000	0.000
log Fundraising Expenses$_{t\text{-}1}$	0.170	0.169	0.164
	0.000	0.000	0.000
log Age$_t$	−0.225	−0.204	−0.204
	0.000	0.000	0.000
log Total Assets$_{t\text{-}1}$	0.468	0.456	0.457
	0.000	0.000	0.000
log Program Service Revenues$_{t\text{-}1}$	−0.032	−0.031	−0.030
	0.000	0.000	0.000
log Other Revenue$_{t\text{-}1}$	0.008	0.008	0.006
	0.005	0.005	0.029
log Government Grants$_{t\text{-}1}$	−0.001	−0.002	−0.002
	0.757	0.515	0.500
Industry fixed effects	Yes	Yes	Yes
N	7,722	7,722	7,722
Adjusted R^2	0.6727	0.6777	0.6784
Model *F*-statistic	1,221.65	1,083.25	1,018.76
	0.000	0.000	0.000
Highest VIF	3.04	3.04	3.08

Notes:

See Table 1 for variable definitions. Bold represents variables of interest.

6. Robustness tests

To confirm our results we conduct several tests of robustness. First, in untabulated results, we included lagged direct donations in our main model which is essentially an examination of whether the change in donations is associated with being rated. If donations and ratings are simultaneously determined, including lagged donations should reduce any observed association between current donations and lagged ratings. In addition, lagged donations

TABLE 9
Robustness—change models

Dependent variable: *ΔDirect Donations*$_t$	(1) Pooled sample Coefficient *p*-value	(2) Matched sample Coefficient *p*-value
Constant	4.409	4.383
	0.000	0.000
New Rating$_{t\text{-}1}$	**3.693**	**0.679**
	0.000	***0.028***
ΔPRICE$_{t\text{-}1}$	−0.009	−3.015
	0.277	0.000
ΔFundraising Expenses$_{t\text{-}1}$	0.082	−0.015
	0.000	0.844
ΔTotal Assets$_{t\text{-}1}$	−0.003	0.099
	0.796	0.014
ΔProgram Service Revenues$_{t\text{-}1}$	−0.009	−0.009
	0.651	0.882
ΔOther Revenue$_{t\text{-}1}$	0.032	0.100
	0.077	0.137
ΔGovernment Grants$_{t\text{-}1}$	0.077	0.276
	0.000	0.000
Industry fixed effects	Yes	Yes
N	12,651	940
Adjusted R^2	0.0234	0.2725
Model *F*-statistic	28.58	32.97
	0.000	0.000
Highest VIF	1.10	1.64

Notes:

See Table 1 for variable definitions. Bold represents variable of interest.

capture the effect of any time-invariant omitted nonprofit-specific factors. After including lagged donations in our model, we continue to find a strong positive relationship between direct donations and our *Rated* test variable.

Our second set of robustness tests are tabulated in Table 7 and exclude hospitals and universities from our sample.[23] As of June 2008, Charity Navigator reported that it is not adding any hospitals, hospital foundations, universities or colleges, land trusts, community foundations, or PBS stations. Therefore, we have reason to believe that these types of nonprofits receive relatively fewer ratings than other nonprofit industries. In addition, theory suggests that nonprofits in which the donor and beneficiary are the same fall under the umbrella of "consumptive philanthropy" whereby donors' decisions to give are relatively insensitive to ratings of performance (Gordon and Khumawala 1999). Specific to our sample, Table 2, panel B indicates that 13 percent of our rated sample comes from the health (including hospitals) industry while over 30 percent of our unrated sample is classified as a health nonprofit. Similarly, education nonprofits make up a mere 7 percent of rated nonprofits in our sample while unrated education nonprofits make up 21 percent of the unrated nonprofits in our sample. Results using this industry restricted sample are consistent with our main results indicating that hospitals and universities are not driving our results.

23. Our sample size in Table 7 is 9,379, from Table 2, panel B: 16,707 less 4,394 health nonprofits less 2,934 educational institutions.

Our third set of robustness tests are tabulated in Table 8. Here, we use a matched sample of nonprofit organizations where we match rated and unrated nonprofits based on nonprofit industry, size (*Total Assets*), and donation reliance (defined as the ratio of direct donations to total revenues) using a propensity score matching procedure.[24] We first match treatment nonprofits with control nonprofits from the same industry and then match to the nonprofit reporting the closest total assets and donation dependency values, without replacement.[25] Results using this matched sample are once again consistent with our main results indicating that nonprofit ratings are significantly associated with more direct donations.[26]

In our final set of analyses we employ change models in an attempt to control for all other factors effecting nonprofit donations. To do so we identify 470 nonprofits that are rated for the first time in 2007 and specify all model variables in change terms. Our dependent variable, *Direct Donations*, is defined as the change in direct donations from 2007 to 2008, and our control variables are defined as the change in values from 2006 to 2007. Table 9 reports the results of this analysis. Column (1) uses a pooled sample of newly rated and unrated nonprofits (we exclude nonprofits that are rated throughout the time period), while column (2) is a matched sample of newly rated and unrated nonprofits.[27] As above, in column (2) we match newly rated (treatment) nonprofits with control nonprofits within the same industry reporting the closest total assets and donation dependency values using propensity score matching, without replacement. In both cases, newly rated nonprofits receive more in contributions relative to unrated nonprofits. These results suggest that the act of being rated, and not self-selection bias, is driving our findings.

7. Conclusion

We find robust evidence which indicates that nonprofit donors incorporate third-party rating information into their decision to give. Specifically, using a sample of over 3,800 unique nonprofit organizations rated by the three largest charity rating organizations in 2007 (Charity Navigator, BBB, and AIP) and over 12,000 unrated control nonprofits, we find that rated nonprofits have significantly higher direct donations than unrated nonprofits. These findings further indicate an incremental increase in direct donations at nonprofits receiving multiple ratings. We also find evidence that receiving a good rating is associated with higher direct donations over nonrated nonprofits, and while poor ratings are also positively related to direct donations, the relationship is not as strong. Finally, we document results indicating that more direct donations accrue to charities that receive consistently good ratings.

Given the time and relative expertise necessary to gather and process nonprofit information in the market for charitable giving, we believe this study makes an important contribution to understanding the role of signals provided by nonprofit rating agencies. Using ratings provided by the three prominent U.S. charity rating agencies as credible signals of charity success, we find support for the applicability of the information-production hypothesis in the market for charitable contributions; concluding that these charitable rating agencies are successful at generating credible signals of quality useful in the donation decision.

24. According to Tucker (2010) researchers face two types of selection bias. Bias from unobservable variables, which should be controlled for using an IMR framework, and observable bias best controlled for using a propensity score matched sample. Given that rating choice is driven by unobservable variables, we have selected an IMR framework for our primary analyses; however, to rule out the possibility that rating choice is actually driven by observable variables, in tests of robustness we also employ a propensity score matched sample design.
25. Our sample size in Table 8 is 7,722; 3,861 rated and 3,861 unrated nonprofits.
26. We also find consistent (untabulated) results when we exclude hospitals and universities from our matched sample in Table 8.
27. Our sample size in Table 9 column 1 is 12,651; our full sample of 16,707 less 4,056 nonprofits rated in both 2006 and 2007. Our sample size in Table 9 column (2) is 470 newly rated nonprofits and 470 unrated propensity score matched control nonprofits.

Appendix

Example of an inconsistent rating: Relief International

Better Business Bureau 2007 (Failed to meet all 20 standards)				
Name of National Charity	Met standards	Standards not met	Unable to verify	Review in progress
Relief International		13, 14	9	

Charity Navigator 2007 Four Star Rating (Greater than 60 points)										
Date published	2011-Apr-01	2010-May-1	2009-Mar-01	2008-Apr-01	2007-Feb-01	2006-May-1	2005-Feb-01	2004-Aug-01	2003-Jun-09	2002-Oct-15
Overall rating	68.65	67.67	63.79	66.24	66.35	66.06	55.25	54.94	45.74	34.69

References

American Institute of Philanthropy. 2012. Available online at http://www.charitywatch.org/, retrieved September 30, 2012.

American Institute of Philanthropy. 2014. Available online at http://www.charitywatch.org/faq.html, retrieved April 12, 2014.

An, H., and K. C. Chan. 2008. Credit ratings and IPO pricing. *Journal of Corporate Finance* 14 (5): 584–95.

Baber, W. R., A. A. Roberts, and G. Visvanathan. 2001. Charitable nonprofits' strategies and program-spending ratios. *Accounting Horizons* 15 (4): 329–43.

Barber, B., R. Lehavy, M. McNichols, and B. Trueman. 2001. Can investors profit from the prophets? Security analyst recommendations and stock returns. *Journal of Finance* 56 (2): 531–63.

Better Business Bureau Wise Giving Alliance. 2012. About BBB wise giving alliance. Available online at http://www.bbb.org/us/Wise-Giving/, retrieved September 30, 2012.

Better Business Bureau Wise Giving Alliance. 2014. Charity complaints and reviews. Available online at http://www.bbb.org/us/Charity-Complaints/, retrieved April 12, 2014.

Bongaerts, D., K. J. M. Cremers, and W. N. Goetzmann. 2012. Tiebreaker: Certification and multiple ratings. *The Journal of Finance* 67 (1): 113–52.

Brennan, M. J., N. Jegadeesh, and B. Swaminathan. 1993. Investment analysis and the adjustment of stock prices to common information. *Review of Financial Studies* 6 (4): 799–824.

Callen, J. L. 1994. Money donations, volunteering and nonprofital efficiency. *The Journal of Productivity Analysis* 5 (3): 215–28.

Charity Navigator. 2012. Available online at http://www.charitynavigator.org/, retrieved September 30, 2012.

Charity Navigator. 2014. Available online at http://www.charitynavigator.org/index.cfm?bay = content.view&cpid = 32#.U0mZZPldV8E, retrieved April 12, 2014.

Chung, W., and A. Kalnins. 2001. Agglomeration effects and performance: A test of the Texas lodging industry. *Strategic Management Journal* 22 (10): 969–88.

Cnaan, R. A., K. Jones, A. Dickin, and M. Salomon. 2011. Nonprofit watchdogs: Do they serve the average donor? *Nonprofit Management and Leadership* 21 (4): 381–97.

Chen, G. 2009. Does meeting standards affect charitable giving? An empirical study of New York metropolitan area charities. *Nonprofit Management and Leadership* 19 (3): 349–65.

Connelly, B. L., T. S. Certo, D. R. Ireland, and C. R. Reutzel. 2011. Signaling theory: A review and assessment. *Journal of Management* 37 (1): 39–67.

Feng, N., Q. Ling, D.G. Neely, and A. Roberts. 2014. Using archival data sources to conduct nonprofit accounting research. *Journal of Public Budgeting, Accounting and Financial Management* 26 (3): 458–93.

Fischer, E., and R. Reuber. 2007. The good, the bad, and the unfamiliar: The challenges of reputation formation facing new firms. *Entrepreneurship Theory and Practice* 31 (1): 53–75.

Gordon, T. P., and S. B. Khumawala. 1999. The demand for not-for-profit financial statements: A model of individual giving. *Journal of Accounting Literature* 18: 31–56.

Gordon, T. P., C. L. Knock, and D. G. Neely. 2009. The role of rating agencies in the market for charitable contributions: An empirical test. *Journal of Accounting and Public Policy* 28 (6): 469–84.

Grant, L. E. 2010. The response to third-party ratings: Evidence of the effects on charitable contributions. Unpublished manuscript, University of California Santa Barbara.

Greenlee, J., and K. Brown. 1999. The impact of accounting information on contributions to charitable nonprofits. *Research in Accounting Regulation* 13: 111–25.

Gujarati, D. N. 1995. *Basic Econometrics*, 3rd edn. New York: McGraw-Hill.

Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153–61.

Independent Sector. 2012a. Scope of the Sector. Available online at http://www.independentsector.org/scope_of_the_sector, retrieved September 6, 2012.

Independent Sector. 2012b. The sector's economic impact. Available online at http://www.independentsector.org/economic_role, retrieved September 6, 2012.

Larcker, D., and T. O. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49 (3): 186–205.

Lennox, C. S., J. R. Francis, and Z. Wang. 2012. Selection models in accounting research. *The Accounting Review* 87 (2): 589–616.

Li, W., E. McDowell, and M. Hu. 2012. Effects of financial efficiency and choice to restrict contributions on individual donations. *Accounting Horizons* 26 (1): 111–23.

Livingston, M., and L. Zhou. 2010. Split bond ratings and information opacity premiums. *Financial Management* 39 (2): 515–32.

Lowell, S., B. Trelstad, and B. Meehan. 2005. The ratings game. *Stanford Social Innovation Review* (Summer): 39–45.

Lundholm, R. J. 1991. Public signals and the equilibrium allocation of private information. *Journal of Accounting Research* 29 (2): 322–41.

Marquardt, D. 1980. You should standardize the predictor variables in your regression models: Discussion of "a critique of some ridge regression methods". *Journal of the American Statistical Association* 75: 87–91.

Milgrom, P. R. 1981. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics* 12 (2): 380–91.

Okten, C., and B. A. Weisbrod. 2000. Determinants of donations in private nonprofit markets. *Journal of Public Economics* 75 (2): 255–72.

Parsons, L. 2003. Is accounting information from nonprofit nonprofits useful to donors? A review of charitable giving and value-relevance. *Journal of Accounting Literature* 22: 104–29.

Parsons, L. 2007. The impact of financial information and voluntary disclosures on contributions to not-for-profit nonprofits. *Behavioral Research in Accounting* 19 (1): 179–96.

Petrovits, C., C. Shakespeare, and A. Shih. 2011. The causes and consequences of internal control problems in nonprofit organizations. *The Accounting Review* 86 (1): 325–57.

Posnett, J., and T. Sandler. 1989. Demand for charity donations in private non-profit markets. *Journal of Public Economics* 40 (2): 184.

Saxton, G., D. G. Neely, and C. Guo. 2014. Web disclosure and the market for charitable contributions. *Journal of Accounting and Public Policy* 33 (2): 127–44.

Skinner, D. J. 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research* 32 (1): 38–60.

Sloan, M. F. 2008. The effects of nonprofit accountability ratings on donor behavior. *Nonprofit and Voluntary Sector Quarterly* 38 (2): 220–36.

Steinberg, R. 1986. The revealed objective functions of nonprofit firms. *Rand Journal of Economics* 17 (4): 508–26.

Tinkelman, D. 1999. Factors affecting the relation between donations to not-for-profit nonprofits and an efficiency ratio. *Research in Government and Nonprofit Accounting* 10: 135–61.

Tinkelman, D. 2004. Using nonprofit nonprofit-level financial data to infer managers' fundraising strategies. *Journal of Public Economics* 88 (9–10): 2181.

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

Tinkelman, D., and D. Neely. 2011. Some econometric issues in studying nonprofit revenue interactions using NCCS data. *Nonprofit and Voluntary Sector Quarterly* 40 (4): 751–61.

Trussel, J. M., and L. M. Parsons. 2007. Financial reporting factors affecting donations to charitable nonprofits. *Advances in Accounting* 23: 263–85.

Tucker, J. W. 2010. Selection bias and econometric remedies in accounting and finance research. *Journal of Accounting Literature* 29: 31–57.

Verrecchia, R. 2001. Essays on disclosure. *Journal of Accounting and Economics* 32 (1–3): 97–180.

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

Yetman, M. H., and R. J. Yetman. 2013. Do donors discount low-quality accounting information? *The Accounting Review* 88 (3): 1041–67.