Better performing NGOs do report more accurately?

Evidence from investigating Ugandan NGO financial accounts*

Dr Ronelle Burger

University of Stellenbosch, South Africa

Canh Thien Dang

University of Nottingham, United Kingdom

Dr Trudy Owens

University of Nottingham, United Kingdom

Corresponding author:

Dr Trudy Owens

School of Economics, Sir Clive Granger Building, The University of Nottingham

University Park, Nottingham, NG7 2RD, U.K.

Email: trudy.owens@nottingham.ac.uk

Tel: +44 (0)115 951 5477

Fax: +44 (0)115 951 4159

Running title: NGO performance and reporting accuracy

* We thank the editor and two anonymous referees for helpful suggestions. We thank Oliver Morrissey for comments on earlier drafts, Julien Labonne, Rafaelle Miniaci, Jan Stuhler for discussions, and participants at OxDev Workshop, German Development Research Group Conference (Gottingen), Nordic Econometrics Meeting (Tartu), Nottingham PhD Conference, North American Summer Meeting of Econometric Society (UC Davis) for comments and feedback. The collection of the Ugandan data was supported by the World Bank and the Japanese Government. Burger thanks the National Research Foundation of South Africa for their support (Grant Number: RCA13102556861). Dang thanks the Economics and Social Science Research Council [grant number ES/J500100/1] for their financial support; the YITP of the IAAE and the University of Brescia (Italy) for their hospitality. Usual disclaimer applies. Conflict of interest: none.

Abstract

Improving ways to assess development outcomes and to identify corruption is considered imperative for NGO sustainability. Yet, little has been done to understand the connection between the effectiveness of NGOs and their financial accountability. We use Benford's Law as a cost-effective and replicable measure to identify fraudulent financial reports of a representative sample of Ugandan NGOs. We find that 25% of the sample provided financial information that did not conform to the Benford distribution, suggesting cases of irregularities that a regulatory body would be advised to investigate. We observe NGOs with better ratings from their beneficiaries are more likely to submit credible financial data. This contradicts the belief that upward accountability demands crowd out serving the client community. We also distinguish between the decision to withhold requested financial information and the decision to report inaccurately. There is no evidence that the two decisions are related, with the decision to not provide all requested financial information attributed to limited capacity and skills. Based on these results, we recommend additional support for NGOs to assist with bookkeeping activities and an expansion of the larger role for beneficiary-based assessments in monitoring the sector.

Keywords: NGOs; misinformation detection; Benford's Law; accountability; selection models

JEL Codes: C21; C52; L31; D82; D23

1. Introduction

Non-governmental organizations (NGOs) have become an integral part of international development efforts. NGOs complement local governments by providing public goods and have become key players in channeling aid from local and international donors. In some cases, NGOs have displaced the traditional role of state institutions when donors, aiming to bypass corrupt governments, deliver aid through these non-state channels (Acht, Mahmoud, and Thiele 2015). Increasingly, decentralized aid directed towards grassroots organisations are viewed as more responsive to the needs of intended beneficiary recipients (Mansuri and Rao 2012).

The sector is not immune from its own cases of scandals related to misuse of funding and falsification of project outcomes (Chen 2016; Herzlinger 1996). Confronted with the continual need for financing, NGOs face growing incentives to produce "rose-colored" project reports that might misrepresent the beneficiaries' circumstances and suppress failures and lessons learnt. This has eroded public trust and increased pressure for a greater level of financial transparency (Ebrahim 2003). The situation is further complicated by the current lack of effective methods to measure transparency and performance of development NGOs (Aldashev and Navarra 2018). Improving ways to identify corruption, assess development effectiveness, and monitor activities has become imperative for the sustainability of the NGO sector.

In this paper, we advocate the use of an accounting technique, Benford's Law, as a scalable and cost-effective tool to measure the level of errors in the financial data reported by development NGOs. We examine the link between self-reported financial data by NGOs and community-based evaluations of their performance, asking whether NGOs with better beneficiary-based assessments also report financial information more accurately. We study the determinants of financial transparency by distinguishing between the decision to withhold financial information when requested and the decision to report the information inaccurately.

Our paper relates to at least three topics of importance for development policy. First, it addresses concerns of development practitioners and donors who are often confronted with challenges in gauging misrepresentation in financial data and red-flagging dubious organizations. The standard measures in the literature, such as program ratios or the distributional discontinuity of specific categories, are of limited use due to scarcity of data (Zitzewitz 2012) and their ability to only detect irregularities in certain items on financial statements for a group of organizations.

We present the first systematic example of using digital analyses to discover information irregularities for individual development NGOs. Using an original and representative dataset of the Ugandan NGO sector (see Barr, Fafchamps, and Owens 2003) and a digital analysis based on Benford's Law, we are able to construct measures of misreporting in financial data for each surveyed Ugandan NGO. We find that 25% of the NGOs provide financial information that potentially differs from their real values.

In a rapidly increasing sector, the method could be a useful screening for targeted auditing and reducing the costs of monitoring. The reason for its scalability and cost effectiveness is it relies on only two pieces of information, namely, the distribution of the leading digits of all numeric figures in an organization's self-reported financial data (the observed distribution) and the theoretical Benford distribution. The method has two underlying principles. First, a dataset that contains naturally occurring numbers (without manipulation) will have the observed distribution of the first digits following a logarithmic distribution. A greater deviation between the observed distribution and the theoretical distribution indicates a greater level of information irregularities. Second, people cannot fabricate datasets that follow the Benford distribution even when instructed to do so (see Schulter, Mittenecker, and Papousek 2010); Boland and Hutchinson 2000). Given the context of a developing NGO in its early days in early 2000s, the Ugandan NGO sector presents a unique opportunity to employ Benford's Law and examine its usefulness for this sector.

The second area of interest for development policy is evaluating NGO performance. Current evaluations of NGOs and projects rely on two sources, namely, NGO self-reported accounts presented in annual statements and community-based evaluations of NGO activities through end-of-project surveys (Aldashev and Navarra 2018). While self-assessed reports are subject to misrepresentation and errors, they are often the only source for evaluating NGOs. Community-based evaluations are difficult to collect due to both logistical challenges and the high cost of surveys. It is therefore important to understand the respective contributions of these two information sources and also the alignment or misalignment between them.

To date, there is no academic work mapping NGO self-reported accounts and beneficiary assessments. A mapping between the two sources provides an opportunity to cross-check NGO self-reported accounts for donors and policymakers who are constrained (logistically or financially) in conducting community surveys. Using performance ratings from a community-based module conducted with the Ugandan NGO survey, we find a robust link between the two sources of NGO evaluation. NGOs that performed better, measured by a higher rate of satisfaction by their respective community, also reported financial information more accurately, measured by the digital analysis based on Benford's Law. A battery of robustness checks shows that our results are not driven by a range of organizational characteristics, potential unobserved within-NGO heterogeneity, or the assumptions of the distributional and functional form.

Finally, we examine the widespread belief in development debates that the high cost of upward accountability requirements such as financial reports crowds out downward accountability, i.e. serving the client community. To do so, we investigate factors influencing two reporting behaviors of NGOs, namely, strategic information withholding and misreporting.

This exercise provides a unique insight into NGO behavior that could be useful when designing aid contracts and incentive schemes to regulate and monitor the sectors financial activities. Indeed, while deliberate non-disclosure of financial accounts remains challenging to address, grassroot development organizations may have a legitimate concern over accountability tasks. NGOs with limited financial and human capital resources are often constrained in expending effort on bookkeeping activities.[1] NGOs could voluntarily overlook some unneeded categories of financial data due to a genuine lack of capacity and staff skills rather than intentional misrepresentation. By comparing different modelling specifications under strategic withholding and genuine missing information, we show missing financial information in the Ugandan NGO sector could result from organizational constraints rather than strategic non-disclosure. The tendency to fully report requested financial information does not correlate with the consequential reporting accuracy. NGOs with more clerks and a larger proportion of staff having a degree have a significantly higher propensity to provide full financial information when requested. Coupled with the evidence of a positive link between performance and reporting accuracy, the prevalence of financial misreporting in the Ugandan NGO sector could be driven by the lack of resources dedicated to accounting tasks rather than manipulation to hide mediocre performance. This result calls for additional support for NGOs to assist with bookkeeping activities.

Our paper contributes to several branches of the literature. First, it adds to the literature on the regulation and activities of NGOs (Hatte and Koenig forthcoming, Aldashev and Navarra 2018). Second, our measure of misreporting relates to the growing literature on forensic economics (Zitzewitz 2012) and the use of distributional properties of numbers to

[1] Epstein and Yuthas (2014) report that 80% of surveyed donors did not provide sufficient overhead allocations to cover the expenses their recipients incurred on reporting requirements.

identify potential misreporting in public data (see Fang and Gong 2017 for detection of overbilling in Medicare reimbursements; and Almond and Xia 2017 for detection of manipulation in investment returns of US non-profits). Our method adds to successful techniques that identify and measure corruption, such as randomized auditing of officials prior to elections (Avis, Ferraz, and Finan 2016) or targeted audits by local governments (Olken 2007). Previous notable and methodological studies using Benford's Law include Amiram, Bozanic, and Rouen (2015) for corporates' financial accounts, Michalski and Stoltz (2013) for national statistics, and Judge and Schechter (2009) for survey data. Like any measure of fraud and corruption detection (see Olken and Pande 2012), digital analysis (e.g. Benford's Law) is not completely fool-proof, nor does it serve as a substitute for auditing, however it does have the advantages of being applicable to any financial and accounting reports and economic.

Third, we examine the underlying mechanism of information disclosure and is the first such paper in development economics, adding to the broader literature on the topic that includes empirical applications in labor economics (Bettin, Lucchetti, and Zazzaro 2012 on remittances) and health economics (Dow and Norton 2003; Madden 2008 on cigarette consumption).

Finally, our study on the link between NGOs' self-reported accounts and community-based evaluations adds to the literature on organizational performance and misreporting behavior of firms (Burns and Kedia 2006) and the recent theoretical papers on the misbehavior and monitoring in the NGO sector (Auriol and Brilon 2018; Aldashev, Jaimovich, and Verdier 2018). Different from these theoretical studies highlighting the strategic interactions within the organizations and between donors and NGOs, we propose support for a separate and novel explanation based on limited resources for irregularities in the development NGO sector, particularly regarding information disclosure behavior.

The paper is organized as follows. Section 2 describes the data. Section 3 describes Benford's Law and its application to our data. Section 4 provides a conceptual framework to motivate our hypothesis, with a focus on distinguishing between selection models. A fuller model is provided in the Online Appendix A. We discuss our econometric strategy and results in Section 5. Section 6 reports a number of robustness checks and Section 7 concludes.

2. The Ugandan NGO data

We match data from two surveys conducted in 2002 with a representative sample of Ugandan NGOs and their beneficiaries. The first survey module (NGO questionnaire) was administered to a random sample of NGOs in 14 districts of Uganda (see Online Appendix Map OA.1) drawn from a verified NGO Registration List held with the Ministry of Planning within the Ugandan Government (details of sampling and a summary of the Ugandan NGO sector are available in (Barr et al. 2003). The face-to-face interviews with an NGO representative (usually the head of the NGO) was conducted by Ugandan field workers.[2] Each session lasted for approximately two hours with the format, length and an overview of the questionnaire communicated to the NGO respondent beforehand. The questionnaire asked 255 questions, including queries regarding the organization's expenditure and income (in 2001 and 2002), funding, governance and activities the NGO undertakes. In brief, the Ugandan NGOs engaged in a range of activities from advocacy (Human rights, HIV/AIDS awareness and prevention), capacity building (Education and Training, Counselling), to support activities to farmers and farming, child-related services, water/sanitation or credit and microfinance provision. On average, the NGOs in our analysis existed for 9.8 years (standard deviation = 9 years) and offered on average four activities. (Barr et al. 2003) provide a comprehensive description of the Ugandan NGO sector.

[2] At the beginning of the interview, the enumerator informed the NGO representative about the survey objective – conducted by academics to dispel confusion and improve knowledge regarding the NGO sector and assured participants that all information collected would be treated confidentially."

Questions regarding revenue and expenditure were designed such that the information was highly aggregated and should have been readily available from a standard annual account. If the respondent was unable to answer the questions, either the NGO representative was asked to request the information from a relevant colleague or the enumerator left the relevant sections of the questionnaire to be completed and collected on a designated day that suited the NGO. For larger NGOs, the enumerator gave the representative the relevant sections before the interview so that the NGO could prepare the required figures. The overriding instruction was to give advance warning of the data needed and ample opportunity to provide the information either directly via the respondent, a relevant colleague if they were unsure about their answers, or a return visit. Table A4 in the Appendix provides a sample of this information. In essence, the financial data collected resembles cash flows and financial statements of for-profit firms. Ideally, we expect a maximum of 60 pieces of financial information (non-zero financial transactions) to be recorded for each organization for each year, making a total of 120 financial items over two years for our digital analysis. In the field, however, we collected much fewer data points for the analysis (43 on average). We address the implications of this in Section 4 and offer several statistical tests to accommodate the smaller sample size.

The second survey module captured community characteristics, needs and perceptions of the NGO via a structured focus group interview conducted with beneficiary communities following a well-defined protocol to ensure comparability across communities. In the initial NGO interview, each NGO was asked to report up to six parishes in which it had been active. One of the parishes was selected at random for a focus group interview.[3] The enumerator contacted the parish leader asking them to recruit between six to ten community members for

[3] The enumerators in the focus group interview used a series of questions to filter out focus groups that did not know about the NGO.

the group meeting. The selection process ensured comparability and consistency of the community evaluation across NGOs. The structured interview assessed how the beneficiary communities evaluated the NGOs surveyed in the first module. The focus group participants were asked to evaluate their satisfaction with the NGO's performance by rating the statement "the people who live in this parish are satisfied with the performance of [NGO]" on a Likert scale (1 = strongly disagree and 5 = strongly agree). We use this satisfaction rating in our main analysis as it captures the general, collective perception of the NGO's performance. In addition, we collected other perceptions of the focus group regarding the NGOs accessibility, competence, importance to the community, and responsiveness to community needs. The Online Appendix G discusses the correlation of these measures with our general measure and the robustness of our results using these alternative ratings. Our conclusions are not altered.[4]

[4] The ideal method to randomly choose focus group participants would have been to randomly choose beneficiaries from a list instead of going through the parish leader. One potential caveat of our selection method is that some of the listed NGOs might have had favorable connections with the parish leaders, who could then have selected focus group participants to give favorable evaluations, thus potentially introducing bias to our study. However, the ideal randomization was not feasible for funding and logistics reasons. In practice, we believe the results are not biased by this concern. Leaders were not told the name of the NGO being evaluated before selecting the meeting participants. It is therefore unlikely that the leaders could have *ex-ante* strategically recruited a biased focus group favoring the NGO. Anecdotally, some enumerators needed to prompt the group about the NGO being evaluated. Barr and Fafchamps (2006) provide further description of the second module.

We match the NGO surveyed in the first module with the respective community from the second module.[5] During the data compiling process, we discovered four enumerators had exhibited "cheating incidents" while conducting the NGO interviews.[6] To ensure that any misreporting detected by our indices are knowingly attributable to the concerned NGO only, we drop the sample collected by these enumerators. We are left with a sample of 104 NGOs matched with 104 communities.

3. Using Benford's Law to detect potential misreporting

There is a well-developed accounting literature focusing on measuring irregularities in financial reports. Methods include accrual-based estimates from the Jones models or analyses using distributional properties of financial figures (see Amiram et al. 2015). There are, however, several weaknesses inherent in both approaches. First, measures estimated from prediction models suffer from sample selection bias and measurement errors as they are based on the error terms of regressions on firms' predicted values (Dechow, Ge, and Schrand 2010). Second, these measures require strong assumptions about the organization's objective function and managers' incentives, which are not always realistic and could induce correlation between the measures and the organization's characteristics. Third, these models require forward-looking information to construct predictions and often detailed time-series and panel data. This requirement often tempers their use in non-profit studies, in which small sample size and comprehensive data collection are the main challenges. The current non-profit literature only focuses on potential errors in some specific categories such as investment returns (Almond and

[5] As there were cases when some NGOs were linked to more than one community, we randomly eliminate 19 duplicates to ensure a 1:1 relationship throughout the analysis.

[6] It is a common problem in collecting survey data (see Judge and Schechter 2009).

Xia 2017), fundraising ratio and program ratios (Hofmann and McSwain 2013). This focus ignores the fact that organizations could manipulate the data as the whole.

We advocate Benford's Law as an alternative way for measuring the accuracy of self-reported financial information when only standard information on a financial statement are available. It is a mathematical law regarding the frequency distribution of leading digits in naturally occurring datasets (e.g., the leading digit of the number 1,201.17 is 1). Contrary to the basic intuition, the occurrence of each digit as the first digit is not equally likely (uniform). Instead, the first digits would follow a decreasing distribution specified in Hill's (1995) theorem. It states that random samples, over different orders of magnitudes taken from a random mix of non-truncated and uncensored distributions, will have the frequency of the first digits converging to the logarithmic of a distribution, dubbed the Benford distribution (See Appendix A).[7] Financial transactions are a popular example of a random mix of non-truncated and uncensored distributions as they arise from interactions and behaviors of independent individuals. Since accurate financial records are based on repeated sums, multiplications, or quotients of prices, quantities, and financial transactions, accurately reported financial and accounting data are often expected to follow Benford's Law (Boyle, 1994). As people are rarely capable of making up datasets that follow a specific distribution (see psychology studies such as Schulter et al. 2010; Hill 1998), biased reports are expected to deviate from the law.

We exploit this property of financial data to uncover information irregularities of Ugandan NGOs. Durtschi, Hillison, and Pacini (2004) outline three requirements for applying Benford's Law to empirical data. First, the data should not have a built-in maximum/minimum. Second, there should not be any externally assigned values. Third, the distribution should be

[7] An order of magnitude is a proxy for the number of the digits that a number has in the base-ten system. For example, the order of magnitude of 14500 is 4 as $14500 = 1.45 \times 10^4$.

positively skewed with a median that is lower than the mean. Our NGO data satisfies all these criteria for applying Benford's Law. The law is also an appealing method for the type of data often available for development NGOs which normally does not allow for the calculation of the other standard measures discussed earlier in this section.[8]

Measuring the extent that a dataset deviates from Benford's distribution has been debated in the digital analysis literature (see Morrow 2014; Miller 2015). Measures can be strongly influenced by the number of digits used, with some statistics requiring near-perfect conformity to the theoretical distribution to not reject the null of conformity (Nigrini 2012). Following Amiram et al. (2015), we use the Mean Absolute Deviation (MAD) statistic in the main analysis, and report results using other popular measures derived from Benford's Law in Appendix B. The MAD statistic is calculated as the mean of the absolute difference between the empirical proportion of each digit in each NGO's aggregated financial reports and their respective theoretical frequency according to Benford's Law (see Table A1):

$$\mathrm{MAD} \equiv \frac{1}{9}\sum_{i=1}^{9}\left|\mathrm{P_o}(\mathrm{d_i}) - \mathrm{P_e}(\mathrm{d_i})\right| \tag{1}$$

[8] There may be concern that our result is driven by random errors introduced during the interview by either the respondent or interviewer rounding up the numbers. This issue is not critical. First, as our measure of information accuracy relies solely on the first digit of the numbers submitted, the usual rounding of the last digit is not a concern. Second, we obtain qualitatively unchanged results when we exclude from our analysis five NGOs whose reported numbers looked as if they were potentially rounded. These NGOs have more than 90% of their submitted financial figures that start by a digit followed by zeros (for example, 8000 shillings). The implication is that heavy rounding does not affect our analysis.

where $\mathrm{d_i} = 1, 2, \ldots, 9$ represents the first digit; $\mathrm{P_o(d_i)}$ is the observed frequency of digit $\mathrm{d_i}$, and $\mathrm{P_e(d_i)}$ is the expected frequency of digit d_i according to Table A1. Using simulated experiments, Amiram et al. (2015) show that the degree of deviation from the Benford distribution strongly correlates with the degree of errors introduced into the financial statements. That is, the larger the MAD statistic, the further the deviation from the theoretical distribution under the null hypothesis that the aggregated report is free of errors.

We construct the degree of reporting accuracy of NGOs by subtracting the level of reporting inaccuracy (MAD) from the (arbitrarily chosen) number one. For NGOs without all requested information, we assign the reporting accuracy as zero to allow for estimating the Cragg model (discussed in the next section). The dependent variable of interest is:[9]

$$\mathrm{R_i} = \begin{cases} 1 - \mathrm{MAD} & \text{if } \mathrm{C_i}(.) = 1 \\ 0 & \text{if } \mathrm{C_i}(.) = 0 \end{cases} \qquad (2)$$

where higher $\mathrm{R_i}$ indicates that the complete report is more accurate.

Since there is no critical value associated with the MAD statistic, it could be impractical for flagging potential misreporting. Due to the small sample size, we employ the two-sample Kolmogorov–Smirnov statistic to test the hypothesis that the observed distribution of all the first digits follows the Benford distribution. Figure 1 presents the results of this exercise. Panel A shows evidence to support the applicability of Benford's Law to our data. When combining available numerical data from all the Ugandan NGOs, the distribution of the first digits of all the financial figures closely follows the Benford distribution, confirming our premise that accounting data follows the law.[10] Based on the Kolmogorov–Smirnov test

---

[9] This assignment does not affect the estimation procedure of Heckman's model as only the sample with complete disclosure is used in the second step.

[10] Note that the conformity of the data does not prevent the possibility that some individual NGOs may have inaccurate financial data. The reason is that the overall conformity may come

statistic, we reject the null hypothesis of complying with Benford's Law for 25% of the NGOs. In the rest of this paper we refer to such cases as the NGOs "deviate" from the law, suggesting that financial accounts of these organizations might frequently contain mistakes, errors or misreporting. Panel B provides a representative distribution of this group. In contrast, Panel C presents a representative NGO out of the remaining 75% of the sample where we fail to reject the null hypothesis. As shorthand we will say that these NGOs "conform" to Benford's law, indicating that we fail to find evidence of misreporting (i.e. cannot reject the hypothesis that the observed distribution follows the Benford distribution).

[Insert Figure 1 here]

4. A conceptual framework

We review previous theoretical papers to conceptualize factors behind reporting behavior of an agent, focusing on whether and how accurately they reveal their financial information. The Online Appendix OA presents a formal model.

4.1 Why do NGOs provide incomplete financial information?

Despite the surveys being designed to make providing financial information easy, several NGOs provided only some of the requested financial figures. In this paper, we define an NGO as providing complete financial information (*complete disclosure*) if their representative provided all requested expenditure-related and revenue-related figures for both 2001 and 2002 fiscal years. There are 77 such NGOs in the sample. We define NGOs as having *incomplete disclosure* if they failed to provide either revenue-related figures or expenditure-related figures or both (those items recorded as missing). The remaining 27 NGOs fall into this category.

---

from a mixture of independent errors embedded in different NGO data. According to Hill's (1995) theorem, these independent errors would result in a mixture of independent distributions whose mixed distribution would follow Benford's Law.

We aim to understand the mechanisms driving these non-disclosure incidents. For NGOs who did not provide all the requested figures (25% of the sample), we must make assumptions about their (potential) accuracy: some of these NGOs may have possessed all the requested financial information but chose (strategically) to provide none or only a part of their data; other NGOs may simply not have had the data (because they did not have the necessary resources or financial skills to compile such information). We want to treat the former as latent observations (i.e. available but not provided) and the latter as actual zeros (the NGO provided all available information, even if the incomplete information yields an accuracy level of zero about their financial situation).

Following Verrecchia's (2001) taxonomy we label the first case discretionary-based disclosure. A sender (here the surveyed NGO) observes private information about the true state of their financial situation and strategically communicates to a receiver (donors or the enumerator) at their own discretion. As such, incomplete disclosure can be optimal due to several *ex-post* costs associated with complete disclosure. First, the information disclosed could reveal either the human capital of the NGO leader, for instance incompetence, or managerial incentives, or the organization's inefficiency (see (Kothari, Shu, and Wysocki 2009). Second, an NGO that revealed unusual expenses might need to exert resources to justify the spending to their beneficiaries or donors. Third, an altruistic NGO who is incentivized to disclose incorrect information may incur costs due to an intrinsic aversion to lying (Gneezy, Rockenbach, and Serra-Garcia 2013). In such cases, the NGO may withhold some financial figures, despite having access to all the requested information. Such strategic incomplete disclosure causes a problem of data observability: the reason underlying partial information may correlate with the latent level of reporting accuracy. Such a correlation could severely bias our estimates of interest.

Again, following Verrechia's (2001), we label the second case as efficiency-based disclosure. Here, non-disclosure is an efficient choice of resource allocation. The intuition is that if the organization commits to disclosure, *ex post* complete disclosure will incur costs of information acquisition, such as hiring professional accountants or spending resources on book-keeping. Expecting that these costs will outweigh any potential benefits, the organization *ex ante* chooses non-disclosure to avoid the costs associated with a full disclosure as a corner solution. In our context, an NGO subject to constrained resources could decide *ex-ante* to gather only information that is necessary and productive to their operation. As such, the decision whether to provide *complete disclosure* is taken before (and independent from) the enumerator's visit and may be separated from the decision on the accuracy of financial information had complete disclosure been acquired. The NGO representative provided all *available* information during the interview even if it was incomplete. In this case, the incomplete disclosure observed is not a strategic communication to withhold information, but rather a corner solution to a maximization problem, subject to some *ex-ante* resource constraint.[11]

4.2 Do NGOs with better performance report more accurately?

We examine whether better performing NGOs, as evaluated by their direct beneficiary community, provide more accurate financial accounts when asked by a third party. One

---

[11] Another explanation is that the NGO has no incentive to reveal their financial situation to the enumerator. The enumerator was instructed to pitch the survey as to enable the Uganda government and donors to better assist NGOs. We, therefore, restrict our analysis based on the two explanations above. To account for potential heterogeneity of the enumerator, such as NGOs responded differently to different enumerators, we include dummies of the enumerators' IDs and receive similar results.

prediction is that high-performance organizations have incentives to disclose more information to differentiate themselves favorably from other organizations with "bad news" and thus avoid the problem of adverse selection (Dye 1986). This reasoning implies that because performance is not directly observable to the interested parties, NGOs who serve their communities well will seek to reveal their performance type: better performing NGOs would have less to hide, thereby provide accurate information. Meanwhile, NGOs who are underperforming would have incentives to manage their figures to increase the possibility of earning new grants, which are often tied to performance (Healy 1985). “Good” organizations also have an incentive to issue detailed and accurate records of their financial situation to avoid potential punishment associated with being detected as untruthful, which could "create doubt about the true motive for which good deeds are performed" (Bénabou and Tirole 2006). We formalize this prediction in the alternative form (the implicit null hypothesis is that there is no significant relationship).

H_{1a}: Higher evaluated performance in serving beneficiary communities is positively associated with the accuracy of the NGO’s financial figures.

Since keeping accurate and up-to-date financial records is costly and time-consuming, as it often requires the NGOs to divert resources away from community services, there may exist a trade-off between community satisfaction and reporting accuracy. That is, an NGO could exert effort in providing better services to their beneficiaries while spending less resources on accountability tasks. We formalize this alternative prediction as:

H_{1b}: Higher evaluated performance in serving beneficiary communities is negatively associated with the accuracy of the NGO’s financial figures.

In sum, there are two explanations for an NGO to provide incomplete information: (i) the NGO strategically withholds some information even when complete information is available; and (ii) the NGO chooses *ex-ante* to record only some necessary information. The mechanism underlying each case is different. While the decision to withhold under (i) may be correlated with the consequent reporting accuracy; the data unavailability under (ii) could be assumed independent of the decision governing the consequent reporting accuracy. We also hypothesize that NGOs with better performance ratings would have an incentive to report credible figures, i.e. closer to their true values. We test this hypothesis against the alternative that NGOs might divert resources away from bookkeeping activities toward actual community services so that better performing organizations fall short of their financial report quality. Figure 2 represents a flowchart to summarize our conceptual framework, highlighting the two cases of incomplete information in our survey.

[Insert Figure 2 here]

5. Econometric methodology and analysis

5.1 Deliberate deviations from Benford's Law and unintended inaccuracies

We first examine whether incomplete information occurs exogenously or endogenously. As reviewed in Section 4, when financial information was requested an NGO could have: first, decided whether to provide the enumerator with all the information requested (both expenditure and revenue related figures); second, if it did provide a full account, the NGO could then have decided on the level of accuracy of the information. Formally, let $C_i(.)$ be a binary function for NGO i such that $C_i(.) = 1$ if i provides all requested information in the first stage, 0 otherwise. Let R_i be the measure of reporting accuracy of NGO i in the second stage once we observe all requested information. X_i, S_i are the observable determinants of the outcome equation (the degree of reporting accuracy) and the selection equation (whether to provide all requested figures). The empirical models can be specified as follows:

$$C_i(.) = \alpha' S_i + v_i \qquad \text{(Stage 1)} \qquad (3)$$

$$R_i^* = \beta' X_i + u_i \qquad \text{(Stage 2)} \qquad (4)$$

where R_i^* is the optimal degree of reporting accuracy of NGO i; v_i, u_i are error terms of the two stages. For NGOs with full information, $R_i^* = R_i$; whereas treating R_i^* for NGOs with incomplete information must depend on the mechanism of why incomplete information arises.

If the incomplete information is an *ex-ante* decision (the efficiency-based disclosure), the two choices on whether to report all requested information, and on how accurately to report, are governed by two independent mechanisms. Incomplete information is a corner solution that could be caused by a budget constraint (gathering full information is *ex-ante* financially infeasible) or the lack of skills or clerks (exogenous sample selection due to independent variables). Formally, once controlling for the observables, there would be no factor unobservable to the econometrician to affect the two choices of the NGO. The error terms in the two stages are uncorrelated $cov(v_i, u_i) = 0$. Incomplete disclosure is a genuine observation and the NGOs with incomplete information could be treated as genuinely having zero-accuracy. The appropriate model would be a Tobit-type with censoring point at $R^* = 0$ as in Equation (5). Practically, we use the Cragg (1971) double-hurdle model for censored responses to allow for two different mechanisms underlying the selection and the outcome decision. Another attractive feature of Cragg's model is that homoscedasticity and normality conditions are less necessary for its consistency (Wooldridge 2010).

$$R_i^* = \begin{cases} R_i & \text{if } C_i(.) > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{(Tobit model)} \qquad (5)$$

In contrast, if the incomplete information is an *ex-post* strategic decision, the two stages are governed by two related mechanisms: NGOs after considering their level of reporting accuracy at Stage 2 would decide on whether to reveal full information at Stage 1 (endogenous sample selection on the dependent variable). Full disclosure is met only when the second-stage optimal level of reporting accuracy exceeds a reservation level of accuracy $\underline{R}$. Otherwise, the NGO strategically withholds some information, causing sample selection bias. Formally, even after controlling for observables, there would still be common factors affecting both the stages (for example, the reservation level of reporting accuracy). The error terms are correlated in this case: $cov(u_i, v_i) \neq 0$. We cannot assign NGOs with incomplete information as having zero-accuracy as before because doing so would lump NGOs genuinely without all requested information and NGOs strategically withholding information. The level of report accuracy of the latter are available but unobservable to the researchers because of the withholding. We assign the value of reporting accuracy for these organizations as a latent observation, instead of genuinely zero. The appropriate model would be a Heckman selection model as:

$$\mathrm{R_i^*} = \begin{cases} \mathrm{R_i} & \text{if } \mathrm{C_i}(.) > 0 \text{ and } \mathrm{R_i^*} \geq \underline{\mathrm{R}} \\ \mathrm{NA} & \text{otherwise} \end{cases} \qquad \text{(Heckman model)} \qquad (6)$$

In sum, we have three categories of NGO: (i) an incomplete disclosure (corner-solution) group who did not keep a complete financial record due to constraints (such as fixed costs of information gathering); (ii) another incomplete disclosure group who kept a complete financial record but withheld some information; and (iii) a group who kept complete financial record which they provided to the enumerators.

To infer the dominating selection mechanism that explains the incomplete disclosure we identify the appropriate model that fits the data statistically, similar to Dow and Norton (2003) and Bettin et al. (2012). The standard approach is to check the significance of the inverse Mills ratio coefficient generated from the Heckman sample selection model. However, Norton, Dow, and Do (2008) and Silva, Tenreyro, and Windmeijer (2015) argue that this does not give reliable information about the ability of the models to describe truncated response data.[12] Instead we use Vuong (1989) non-nested hypothesis LR test (see Online Appendix B) and a regression-based specification test under heteroskedasticity developed by (Silva et al. 2015)

[12] Dow and Norton (2003) instead propose an adjusted empirical mean square error test, which is computationally more difficult to implement.

based on Davidson and MacKinnon (1981).[13] If the Cragg's model is preferred, we conclude that once controlling for NGO characteristics, the decision to have incomplete information is unrelated to the decision to manipulate the latent report. Otherwise, if the Heckman's model is preferred this indicates strategic withholding of information. To our knowledge, this approach is the first conducted in the literature on information asymmetry.

5.2 The empirical specification

Following the hypothesis in Section 4, the estimation equations for the degree of accuracy and the selection model for providing all requested information are: [14]

$$R_i^* = \alpha_1 + \gamma \text{Performance}_\text{i} + X_i'\beta_1 + u_i \tag{7}$$

$$C_i(.) = \alpha_2 + \lambda \text{Performance}_\text{i} + X_i'\beta_2 + Z_{2i}\gamma_2 + v_i \tag{8}$$

where R_i is the accuracy measure from Benford's Law, and $Performance_i$ is the evaluation of NGO *'s* performance as measured by the respective community(ranked on the Likert scale: 1 = least satisfied and 5 = most satisfied).

Z_{2i} is the exclusion restriction variable to ensure the consistency of the Heckman estimation. We use a binary variable of whether members of the NGO need to vote before the organization introduces any new activity (*Member involvement* = 1 if yes, 0 otherwise). The rationale for the exclusion restriction is as follows. Stronger member involvement may create incentives and pressure for the NGO to have financial records ready for inspection. Yet, there

13 We implement the test using command –hpc-- in Stata, provided by Silva et al. (2015).

14 Within-district observations might share common characteristics (a local social norm) that may affect the decisions regarding disclosure and accuracy. Including dummies for 14 districts is not a viable option as it substantially reduces the degree of freedom. We report here robust standard errors clustered at the district level (acknowledging that 14 districts are too few for reliable clustering) as the results are qualitatively similar without clustering.

is no reason to believe such involvement would directly impact the accuracy of financial records that the NGO prepares. Olken (2007) uses randomized control trials to show that grassroots participations of the beneficiary communities in monitoring projects had little impact on corrupt behavior of the contracted agents. We expect that *Member involvement* is a significant explanatory variable of the selection mechanism but can be excluded from the equation of reporting accuracy.

X_i is a vector of NGO characteristics suggested in the existing literature as possible explanatory variables of information disclosure and reporting accuracy of nonprofit organizations. To account for the concern that smaller NGOs may have fewer transactions that make their data inherently noisier, we include *Number of non-zeros* to capture the number of non-zero financial transactions used to calculate our indices by Benford's Law. The rest of the controls are as follows.

Monitoring body: We include two variables: (i) whether the NGO has a board of directors or trustees to oversee its activities (*Oversight Board* = 1 if yes, 0 otherwise) and (ii) whether the NGO has officially registered as a company (*Registered as Company* = 1 if yes, 0 otherwise) to capture the effect of company status on NGO behavior by placing more scrutiny and regulations on registered companies.

Financing source: We proxy for the accountability pressure from donors by a binary of whether the NGO ever received a grant (*Received Grant* = 1 if yes, 0 otherwise).

Cost of information gathering: We capture the resources needed for accounting procedures with: (i) the number of clerical staff working for the NGOs (*Clerical staff)* and (ii) the proportion of paid employees having a tertiary education or a degree (% *Professional Degrees*). We use: (i) *NGO Age* (the years of existence to 2002) to proxy for reputation concern and $NGO\ Age^2$ to account for potential differences in behavior of the more established NGOs, and (ii) the number of years the manager has been with the NGO (*Tenure*) to examine the effect of career concerns over financial misreporting (Fudenberg and Tirole 1995). Following Burger and Owens (2010), we use the number of reports per year required by granting bodies (*Reports per year*) to capture reporting fatigue due to heavier administrative costs.

Religious adherence: a binary variable of whether the NGO is affiliated with a religion (*Religion* = 1 if yes, 0 otherwise). One mechanism whereby religion might influence organizational transparency is through the frequent reminder of (religious) moral codes of conduct (McGuire, Omer, and Sharp 2012).

Attitude towards governance. First, we include two binary variables for whether the NGO cites *Lack of funding* (= 1 if yes, 0 otherwise) and *Lack of skilled staff* (= 1 if yes, 0 otherwise) as constraints preventing them from doing an even better job. Second, we use a binary variable of whether the NGO views the local government as a hindrance to their activities (*Government as hindrance* = 1 if yes, 0 otherwise) to capture the antagonistic idealism that may oppose the sectoral norm of accountability and donor demands publicly. Third, we use a continuous variable of how long the manager worked in the Ugandan government before joining the NGO (*Years working in government*) to capture both political connection of these NGOs and potential spill-over effects of the corruption in Ugandan government and public service departments in the 2000s (Deininger and Mpuga 2005).

5.3 Descriptive statistics

Table 1 provides descriptive statistics of the data. 77 NGOs (74% of the sample) provided information on both revenue and expenses information. There is no significant difference in performance between these 77 NGOs and the 27 NGOs that did not provide complete data. Of those that provided information, 58 conform to the Benford distribution of first digits using the two-sample Kolmogorov–Smirnov test at the 10% confidence level. The average performance of these NGOs was higher than that of the 19 NGOs whose financial figures did not conform. To preview, Figure 3 supports our hypothesis of a positive mapping between the evaluated performance and the level of report accuracy when using the naïve OLS estimation without correcting for potential sample selection bias and missing information. We report the estimates in Column (3) in Table OA3 (the Online Appendix). We find a positive relationship between the measure of reporting accuracy and each of the alternative measures of NGO performance, but estimates are not as precise.

[Insert Table 1 here]

[Insert Figure 3 here]

5.4 Complete disclosure and determinants

Table 2 presents the results for determinants of the NGO's decision to provide an incomplete set of financial figures even though the requested information was standard and should have been readily available. The estimates are as expected. Having an oversight board of directors or trustees is significantly and negatively associated with the propensity to provide all requested information to a third party. One explanation is the "unintended chilling effect" suggested by Cormier, Magnan, and Van Velthoven (2005). That is, for established organizations, the board of directors places less pressure on internal governance. In our case, the oversight board may place lesser importance on a full record of standard financial figures, and the NGO, subject to board agreement, may have decided to keep only necessary data. Being registered as a company is not significantly associated with a higher propensity to complete disclosure. This result raises a question of effective monitoring over companies by responsible ministers in the Ugandan government office in the early 2000s (for a similar discussion, see Deininger and Mpuga 2005).

Having received a grant is a positive predictor of complete disclosure. Previous studies posit that organizations that need to facilitate fundraising activities have a higher propensity for information disclosure (Jensen and Meckling 1979). Transparent financial records may serve as a signaling vehicle to attract funding. Having a religious affiliation is a positive predictor of complete disclosure, replicating McGuire et al (2011). Regular reminders of moral codes through religious preaching could be effective in promoting transparency.

We also find significant associations of reputation loss and manager's career concerns with the propensity to complete disclosure. There exists a significant U-shape effect of the organization's reputation proxied by *NGO Age* and *NGO Age*^{2}. More established NGOs tend to conform to the sectoral norm in financial reporting; while older organizations may become complacent over time and be more relaxed with the sectoral standard (the "unintended chilling effect"). Regarding career concerns, there exists a significantly negative relationship between the manager's tenure length and the disclosure propensity. Besides the chilling effect, senior managers may care less about the future job market which commonly values transparency. These managers are also the most familiar with the operations of the NGO and may decide that selectively keeping necessary information may be an efficient strategy.

There is no statistical association between poorer performance and incomplete disclosure, suggesting that it is not the performance that incentivizes the NGOs to withhold information from a third party. Instead, we find evidence that lack of human resources could be a significant constraint towards transparency in the Ugandan NGO sector. NGOs with fewer clerks and a smaller proportion of staff holding degrees are significantly less likely to provide all requested information. Organizations endowed with a smaller qualified workforce may be constrained and possibly reluctant to expend resources to keep all the standard financial data.

NGOs whose managers consider government as a hindrance to their daily operation are less likely to provide all requested financial information. One potential explanation is that they perceive the government as antagonistic and may worry about how the government would use their organizational information. This would capture service delivery NGOs that are critical towards the government and organizations with a political advocacy or watchdog role. Managers may have responded to the lack of human resources and transparency burden by selectively recording figures they saw as necessary to the NGO's operation. We also find that enabling members to vote for new services is strongly correlated with the higher propensity to complete disclosure. This result highlights the benefit of involving members to pressurize management on transparency and accountability issues. The significance of *Member Involvement* also gives support for the exclusion restriction of our Heckman estimation.

There are several surprising results. We find no significant association between the number of reports requested per year by granting bodies and the propensity to provide all standard information. Nor do we find NGOs who claim lack of skilled staff and funding as major constraints to their operation to have a different propensity to provide all requested information. There are two explanations for the insignificant results. First, although the signs of the variables are as expected, the small sample size may lead to imprecise estimates. Second, there may be biases due to endogeneity. Table OA1 in the Online Appendix shows that claiming lack of skilled staff as a major constraint becomes significantly associated with a lower propensity to be transparent once we control for endogeneity (see Section 6.1).

[Insert Table 2 here]

Table 3 presents the results for Vuong's (1989) and the HPC by Silva et al. (2015) model selection tests to distinguish the mechanism underlying the data unavailability.[15] Both the statistics unanimously indicate the double-hurdle model fits our data better than the Heckman model. Statistically, the mechanism underlying *incomplete disclosure* is a corner-solution censoring – controlling for observables, any unobservable confounders underlying the decision to complete disclosure and the decision of report accuracy, are uncorrelated. We reject

[15] According to Silva et al. (2015), the name stands for a test under heteroskedasticity (*H*), combining the *P* test, which conditions on estimates under the alternative, and the *C* test, which conditions on estimates under both the null and the alternative.

the hypothesis that there exists some reservation level of report accuracy such that the NGO would withhold some financial figures if their report accuracy fails to exceed the reservation level. Our discussion above suggests the fixed costs of information gathering and human resource constraints being the key reasons for providing incomplete financial information.

Although a corner solution censoring mechanism fits the Ugandan data better, the two models provide similar results. We refer to the estimates using the double-hurdle model in the next section, noting that the discussion is also valid for the Heckman model.

[Insert Table 3 here]

5.5 Mapping report accuracy and NGO performance

We support the hypothesis that NGOs with higher evaluated performance report financial information more accurately, regardless of the mechanism underlying their disclosure policy. The evaluated performance, however, is an insignificant predictor of the propensity to provide all requested financial data. We interpret the insignificance of performance as further support for a double-hurdle censoring model: two independent mechanisms govern the two decisions. The decision to disclose all requested information is not related to the performance of the organization but rather the NGOs capacity and human capital. The decision to report accurately, however, relates to higher ratings of performance, perhaps to reveal their type or avoid punishment of being detected as untruthful as in Benabou and Tirole (2006).

The estimates for other NGO characteristics in Table 2 align with our expectations. Having registered as a company positively correlates with higher reporting accuracy, possibly due to regulatory effects or interactions with other companies that are more experienced in accounting tasks. A U-shape relationship also emerges between the organization age and the accuracy of their reported data. While more established NGOs report more accurately, the data for the most established organizations becomes less credible. Besides the complacency explanation, the scale of bookkeeping in larger organizations may lead to more errors in the process of information gathering if the clerical staff in charge do not receive sufficient training (Keating and Frumkin 2003). We do not find any significant association of having more clerks and staff with professional degrees on reporting accuracy. Although having a larger workforce increases the propensity to complete disclosure, it is not statistically associated with higher report accuracy. Consistent with anecdotal observations in the Ugandan NGO sector, we conjecture that the lack of training may be one explanation. Another supporting piece of evidence is that a higher number of reported financial figures (*Number of non-zeros*) positively correlates with the conformity of the data with the Benford distribution.

Longer exposure to government bureaucracy correlates with a lower degree of report accuracy. One explanation is the spill-over effect of the corruption prevalent in Ugandan government offices during the time (McCormick 1990). Since the Ugandan public sector in early 2000s was plagued with corruption (Deininger and Mpuga 2005), NGO managers switching from government jobs may have carried the ethos to their new positions. This is in line with Chaney, Faccio, and Parsley (2011) who found that politically connected firms reported poorer quality accounting information than non-connected firms. The political connection reduces the need to respond to regulatory pressures from the authority and the donation market, thus allowing them to disclose lower quality accounting information. Our results highlight the need to strengthen the accountability of the overseeing body to effectively monitor the NGO sector.

Regarding reporting fatigue, more burdensome reporting requests from donors are significantly associated with lower reporting accuracy. Consistent with Burger and Owens' (2010) finding on information misrepresentation, NGOs could submit lower quality financial data as a useful deflection strategy in response to heavy, and possibly unreasonable, demands from donors, while complying sufficiently to maintain grants. Summary statistics in Table 1 show that NGOs with lower report quality are requested to submit on average two reports per year (and those with the lowest accuracy can have up to 12 reports requested each year).[16]

6. Robustness checks

We next report the robustness of our results to potential endogeneity between performance and reporting accuracy, alternative measures of reporting accuracy, and the functional form and distribution assumptions of the specifications.

6.1 Robust to endogeneity between performance and reporting accuracy

The main analysis accounts for sample selection by a Heckman correction model. Yet, there remains another concern of omitted variable bias, namely, altruistic NGOs could self-commit to deliver both higher performance and financial transparency. Although we aim to mitigate the bias caused by the omitted self-commitment with a battery of control variables, the concern remains possible. We propose an instrumental variable strategy to reduce the bias.

A valid set of instrumental variables, Z_i, should have a strong correlation with the evaluated performance but be independent of any strategic behavior of the concerned NGO. Finding the set is challenging for two reasons. First, NGO characteristics are unlikely to satisfy the exclusion restrictions as they could be correlated with the unobserved NGO ability. Second, most of the community characteristics (e.g. available infrastructure, prosperity indicators,

---

[16] One counterargument is that donors simply adjust their demands towards organizations that are more likely to behave in a dubious manner. This unobservable heterogeneity could bias our finding. We argue our explanation still holds. First, as the Ugandan NGO sector had been expanding both horizontally and vertically, there was hardly a shortage of organizations available so that the donors had to compromise and use dubious NGOs. Second, even after controlling for the unobserved heterogeneity that relates to the organization’s ability and potential manipulation (see Section 6.1), the negative association remains (see Table A3).

employment rates) are also likely to be invalid as more able NGOs could strategically locate in convenient areas that could enable them to serve the community better (Brass 2012). We propose two instruments from the characteristics of the focus group participants: (i) the percentage of the group older than 55 and (ii) the percentage of the group who have a connection with the concerned NGO, e.g. a staff or an NGO member.[17]

Our identification assumption is that certain demographics of the focus group (connection with the NGO and average age of focus group members) could strongly predict the (evaluated) performance of an NGO. Groups with more respondents with a connection to an NGO may report higher satisfaction because the NGO has exerted effort in addressing specific needs of the community. Similarly, if the needs of a particular demographic are less well-served we would expect a focus group with a higher proportion of this demographic to result in lower satisfaction scores. The intuition comes from documented institutional characteristics of the NGO sector. Barr and Fafchamps (2006) and Fruttero and Gauri (2005) show that many NGOs are often less inclined to work in remote and poorer communities despite these places often being the neediest and requiring the most attention. Evidence suggests some NGOs do not always try their best to serve the community, resulting in heterogeneous degrees of communities' satisfaction. Likewise, specific to Uganda there is evidence that no NGO reported any service or activity that specifically targeted the elderly (Barr et al. 2003). Using this intuition, we expect that groups with more respondents with a connection to an NGO would report higher satisfaction because the NGO has exerted effort in interacting with their community. In addition, we expect NGOs working in areas with a larger proportion of elderly would receive a significantly lower satisfaction score since the needs of

[17] In the survey beneficiaries were asked to report the number of focus group participants between the age of under 25; from 25 to 39; from 40 to 54 and over 55.

this group are less well-served. F-tests and coefficients in Panel B in Table 4 confirm our conjectures. Groups with more participants having a connection with the evaluated NGO tend to give higher scores; whereas groups with more senior-age members give unfavorable assessments – significant at the 5 and 1 percent level, respectively.

We believe the instruments are exogenous to the accuracy of the financial report given to a third party for three reasons. First, there is no reason why demographics of the community, like age, could affect the tendency of an NGO to report accurately and responsibly to a third party. The demographics could affect how transparent the NGO is towards the village, but hardly with a third party (here the enumerator). Second, as explained in Section 2, the focus group selection can be considered exogenous to the concerned NGO. The community leader selected the participants without knowing the NGO that was to be evaluated. It was therefore not possible for the leader to choose a focus group that would give a biased evaluation of a specific NGO. Indeed, Table A3 shows the instruments do not exhibit any significant correlations with observable characteristics of the NGO working in the community. The balanced tests show that no characteristics of the NGOs are predictive of the composition of the focus group, suggesting the characteristics are as good as random. Third, even if the leader failed to unbiasedly form the focus groups, the NGOs would not possibly self-select into villages by using the demographic statistics used as our instruments. Unlike other community characteristics relating to infrastructure and prosperity or the presence of other public services (see Barr and Fafchamps 2006 for Uganda; Fruttero and Gauri 2005 for Bangladesh), our instruments are unlikely to be a priority in location choice of these Ugandan NGOs. Panel B in Table 4 provides Hansen J statistics and Anderson-Rubin Wald test for the null hypothesis that the orthogonality conditions are valid. In both tests, we fail to reject the null hypothesis that our instruments can be excluded from the main equation.

Panel A in Table 4 reports the results using the Cragg's and Heckman model with endogenous regressors and the proposed instruments for $Performance_i$ (see Online Appendix C for our computational procedures). Both Cragg-Donald (4.485) and Kleibergen-Paap (8.920) statistics are higher than the usual threshold 4, suggesting our instrumentation being informative. To account for the low first-stage F-statistics, we additionally report Anderson-Rubin confidence intervals in Table 4. The confidence interval is [0.6, 60], further supporting the significantly positive relationship between community-evaluated performance and reporting accuracy in Panel A. The implication is that we can rule out the confounding effect of any unobserved motivations that might drive the positive link between better performance in serving the communities and the accuracy of NGO financial information.

[Insert Table 4 here]

6.2 Robust to alternative measures of reporting accuracy

We show that the positive relationship between performance and reporting accuracy is not driven by the method of measuring the deviations from Benford's Law. Appendix B presents three other "critical-value" based proxies used in the Benford' Law literature, namely the Cho-Gaines' (2007) *d*-statistics, the Kolmogorov – Smirnov (KS) statistics, and the KS test for fit of distributions. Although these methods require different choices of critical values, which is prone to subjectivity, they offer ease of use and practical interpretations (Barabesi et al. 2018). Table A2 reports the qualitatively unchanged results for the alternative measures.

6.3 Robustness to the functional form and distributional assumptions of the error terms

Since Heckman and Cragg's estimates rely on the assumptions of normality and heteroscedasticity in the main equation error terms, we report three tests of normality and heteroscedasticity. Figure A1 in Appendix C reports a clear graphical resemblance between the predicted residual density and the normal distribution. In Table 2, Jaque-Bera test statistics strongly supports residual normality, while we barely fail to reject the null hypothesis of homoscedasticity at 5%. In the Online Appendix D, we demonstrate the robustness of the positive mapping between the community-based evaluation and reporting accuracy. In brief, we first estimate two simple nonparametric kernel regressions (local polynomial smooth and lowess smother) for a bivariate relationship between performance and reporting accuracy. The nonparametric estimators relax the functional form assumptions in the main analysis. Second, we perform a Robinson (1988) semiparametric estimator for the sample selection model and a Powell (1984) censored least absolute deviations (CLAD) for the censored Tobit model. The estimators are robust to heteroscedasticity, consistent, and asymptotically normal for a wide class of error distributions. All of the estimations point toward a significantly positive relationship between the two variables, assuring us that our results are not sensitive to the distributional and functional-form assumptions made under the Heckman and Cragg's model.

7. Concluding remarks

In this paper, we demonstrate the feasibility and usefulness of Benford’s Law to study irregularities in financial accounts of a representative sample of NGOs in Uganda. We find that 25% of the sample provided financial information that did not conform to the Benford distribution, suggesting cases of irregularities that a regulatory body would be advised to investigate. Our method allows for a cost-effective and replicable measure to monitor financial records and identify corruption. Like any measure of financial manipulation, our method is neither definitive nor completely fool-proof. Nevertheless, we believe the method is a practical approach to flagging potential misreporting by organizations that could subsequently be targeted for auditing. As suggested by psychology studies, any creative manipulation to preserve the Benford distribution is highly unlikely, especially in the development NGO context when the use of Benford’s law is not well-known.

We also find that the underlying mechanism for nondisclosure of financial accounts and the provision of inaccurate information to be uncorrelated. There is evidence that the shortage of skills and resources contributed to non-disclosure of Ugandan NGOs. Given the fiscal constraints of government-funded regulators in developing countries, it may be necessary to mobilize the donor community to contribute towards funding such oversight mechanisms. This funding mobilization could help harmonize the accountability pressure on organizations with limited-resources, especially when under-staffed NGOs are unable to accommodate both community services and expensive bookkeeping procedures. As it becomes more difficult for NGOs that fail to provide credible financial reports to receive funding in the future, the limited resources devoted to accountability tasks could lead to a vicious cycle within the NGO field. Over time, oligopolization of the NGO sector could become a serious problem where donors could move away from under-staffed organizations due to concerns over their financial transparency.[18] Designing specific financial packages to support poorly resourced NGOs with their bookkeeping activities could be a first step to improve the accountability of the sector and still maintain the benefits that come with grassroots organizations.

The analysis also shows that higher community satisfaction scores are aligned with accurate reporting, challenging the widely held belief that upward and downward accountability are in conflict. Given the alignment and the increased emphasis on community responsiveness and community assessments, this work provides support for the prioritization of independent community-based feedback and assessment sessions over the emphasis on onerous and frequent financial reporting expected in different donor reporting templates. Instead of increasing reporting requirements, which typically requires scarce resources diverted away from the organization's main charitable activities, we provide evidence that collecting assessments from the beneficiary communities may be more efficient. We also find a strong correlation between excessive reporting requirements and lower levels of accuracy. The

[18] We thank an anonymous referee for their excellent input on this point.

implication is that when a reasonable reporting burden is exceeded, cynicism may erode the commitment to accurate and transparent reporting of organizations with limited resources.

Although our study may suffer from small sample bias, the results are robust across a range of modelling approaches: potential endogeneity or relaxing distributional and functional form assumptions. Given the importance of transparency for good governance in the NGO sector and the lack of data and evidence in the literature, further research is vital. Data availability remains an important constraint and limits the evidence available on how to best empower, enable and support this development sector.

Appendix

A. The Benford's distribution of first digits of numbers in a naturally occurred dataset

Hill's (1995) theorem also provides the following formal derivation of the distribution according to Benford's Law:

$$P(d) = \log_{10}(1 + \frac{1}{d}) \tag{A1}$$

where P(d) is the probability that digits $d = 1, 2, \ldots, 9$ occurs as the leading digit in a naturally drawn set of numbers. Table A1 records the full theoretical distribution specified by Benford's Law.

[Insert Table A1 here]

B. Robustness to alternative measures of conformity to Benford's Law

We complement the main analysis with three "critical-value based" measures created from: (1) the Cho-Gaines' (2007) *d*-statistics (*D*), (2) the Kolmogorov – Smirnov (*KS*) statistics, and (3) a binary variable of whether we fail to reject the null hypothesis of the data conforming to the Benford distribution using the two-sample Kolmogorov – Smirnov test at the significance of 10% (*Conform* = 1 if Yes, 0 otherwise):

$$D \equiv 5 - \left[\sum_{i=1}^{9} [P_o(d_i) - P_e(d_i)]^2\right]^{\frac{1}{2}} \tag{A2}$$

$$\text{KS} \equiv 1 - \max_{d_i \in \{1,2,\ldots,9\}} \left| \sum_{i=1}^{d_i} [P_o(d_i) - P_e(d_i)] \right| \tag{A3}$$

$$Conform = \begin{cases} 1 & \text{if } KS > \dfrac{c(\alpha)}{\sqrt{2N}} \\ 0 & \text{if } KS \leq \dfrac{c(\alpha)}{\sqrt{2N}} \end{cases} \tag{A4}$$

where N is the total number of non-zero financial items used, $P_o(d_i)$ and $P_e(d_i)$ are respectively the frequency of digit d_i appearing in the observed sample and the expected (theoretical) distribution, $c(\alpha) = \sqrt{-\frac{1}{2}\ln(\frac{\alpha}{2})}$ is the critical value of the two-sample Kolmogorov – Smirnov test at N and test power $\alpha = 0.10$ (for which $c(\alpha) = 1.22$). For the Cho-Gaines' (2007) d-statistics, we use 5 instead of 1 to construct the measures based on Cho-Gaines' d-statistics since the statistics are larger than 1 (see Morrow, 2014 for the critical values). The number 5 is arbitrarily chosen to facilitate the computation of the Cragg model and varying its magnitude does not change the results. We also transform the usual KS statistics (that measures the divergence of the two samples) into the formula above to measure the convergence of the two samples. Like the measure from the MAD statistic, lower values of the indices indicate that the tested data diverge further from the Benford distribution. Table A2 show similar results for the main equation.

[Insert Table A2 here]

C. Additional Figures

[Insert Figure A1 here]

[Insert Table A3 here]

[Insert Table A4 here]

References

Acht, Martin, Toman Omar Mahmoud, and Rainer Thiele. 2015. “Corrupt Governments Do Not Receive More State-to-State Aid: Governance and the Delivery of Foreign Aid through Non-State Actors.” *Journal of Development Economics* 114 (May). North-Holland: 20–33. doi:10.1016/J.JDEVECO.2014.11.005.

Aldashev, Gani, Esteban Jaimovich, and Thierry Verdier. 2018. “Small Is Beautiful: Motivational Allocation in the Nonprofit Sector.” *Journal of the European Economic Association* 16 (3). Oxford University Press: 730–80. doi:10.1093/jeea/jvx024.

Aldashev, Gani, and Cecilia Navarra. 2018. “Development NGOs: Basic Facts.” *Annals of Public and Cooperative Economics* 89 (1). Wiley/Blackwell (10.1111): 125–55. doi:10.1111/apce.12188.

Almond, Douglas, and Xing Xia. 2017. “Do Nonprofits Manipulate Investment Returns?” *Economics Letters* 155 (June). North-Holland: 62–66. doi:10.1016/j.econlet.2017.03.014.

Amiram, Dan, Zahn Bozanic, and Ethan Rouen. 2015. “Financial Statement Errors: Evidence from the Distributional Properties of Financial Statement Numbers.” *Review of Accounting Studies* 20 (4). Springer US: 1540–93. doi:10.1007/s11142-015-9333-z.

Auriol, Emmanuelle, and Stefanie Brilon. 2018. “Nonprofits in the Field : An Economic Analysis of Peer Monitoring and Sabotage.” *Annals of Public and Cooperative Economics* 89 (1). Wiley/Blackwell (10.1111): 157–74. doi:10.1111/apce.12197.

Avis, Eric, Claudio Ferraz, and Frederico Finan. 2016. “Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians.” Cambridge, MA. doi:10.3386/w22443.

Barabesi, Lucio, Andrea Cerasa, Andrea Cerioli, and Domenico Perrotta. 2018. “Goodness-of-Fit Testing for the Newcomb-Benford Law With Application to the Detection of Customs Fraud.” *Journal of Business and Economic Statistics* 36 (2). Taylor & Francis:

346–58. doi:10.1080/07350015.2016.1172014.

Barr, Abigail, and Marcel Fafchamps. 2006. “A Client-Community Assessment of the NGO Sector in Uganda.” *Journal of Development Studies* 42 (4). Routledge: 611–39. doi:10.1080/00220380600682009.

Barr, Abigail, Marcel Fafchamps, and Trudy Owens. 2003. “Non-Governmental Organisations in Uganda: A Report to the Government of Uganda.” https://pdfs.semanticscholar.org/dffc/c4b5ca10454f7e3588959ec28aeea51b58a3.pdf.

Bénabou, Roland, and Jean Tirole. 2006. “Incentives and Prosocial Behavior.” *American Economic Review* 96 (5): 1652–78. doi:10.1257/aer.96.5.1652.

Bettin, Giulia, Riccardo Lucchetti, and Alberto Zazzaro. 2012. “Endogeneity and Sample Selection in a Model for Remittances.” *Journal of Development Economics* 99 (2). Elsevier B.V.: 370–84. doi:10.1016/j.jdeveco.2012.05.001.

Boland, Philip J., and Kevin Hutchinson. 2000. “Student Selection of Random Digits.” *Journal of the Royal Statistical Society Series D: The Statistician* 49 (4). Wiley/Blackwell (10.1111): 519–29. doi:10.1111/1467-9884.00250.

Brass, Jennifer N. 2012. “Why Do NGOs Go Where They Go? Evidence from Kenya.” *World Development* 40 (2). Pergamon: 387–401. doi:10.1016/J.WORLDDEV.2011.07.017.

Burger, Ronelle, and Trudy Owens. 2010. “Promoting Transparency in the NGO Sector: Examining the Availability and Reliability of Self-Reported Data.” *World Development* 38 (9). Pergamon: 1263–77. doi:10.1016/j.worlddev.2009.12.018.

Burns, Natasha, and Simi Kedia. 2006. “The Impact of Performance-Based Compensation on Misreporting.” *Journal of Financial Economics* 79 (1). North-Holland: 35–67. doi:10.1016/j.jfineco.2004.12.003.

Chaney, Paul K., Mara Faccio, and David Parsley. 2011. “The Quality of Accounting

Information in Politically Connected Firms." *Journal of Accounting and Economics* 51 (1–2). North-Holland: 58–76. doi:10.1016/j.jacceco.2010.07.003.

Chen, Qiu. 2016. "Director Monitoring of Expense Misreporting in Nonprofit Organizations: The Effects of Expense Disclosure Transparency, Donor Evaluation Focus and Organization Performance." *Contemporary Accounting Research* 33 (4). Wiley/Blackwell (10.1111): 1601–24. doi:10.1111/1911-3846.12218.

Cho, Wendy K.Tam, and Brian J Gaines. 2007. "Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance." *American Statistician* 61 (3). Taylor & Francis: 218–23. doi:10.1198/000313007X223496.

Cormier, Denis, Michel Magnan, and Barbara Van Velthoven. 2005. "Environmental Disclosure Quality in Large German Companies: Economic Incentives, Public Pressures or Institutional Conditions?" *European Accounting Review* 14 (1). Routledge Journals: 3–39. doi:10.1080/0963818042000339617.

Cragg, John G. 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica* 39 (5). The Econometric Society: 829. doi:10.2307/1909582.

Davidson, Russell, and James G. MacKinnon. 1981. "Several Tests for Model Specification in the Presence of Alternative Hypotheses." *Econometrica* 49 (3). The Econometric Society: 781–93. doi:10.2307/1911522.

Dechow, Patricia, Weili Ge, and Catherine Schrand. 2010. "Understanding Earnings Quality: A Review of the Proxies, Their Determinants and Their Consequences." *Journal of Accounting and Economics* 50 (2–3). North-Holland: 344–401. doi:10.1016/j.jacceco.2010.09.001.

Deininger, Klaus, and Paul Mpuga. 2005. "Does Greater Accountability Improve the Quality of Public Service Delivery? Evidence from Uganda." *World Development* 33 (1).

Pergamon: 171–91. doi:10.1016/j.worlddev.2004.09.002.

Dow, William H., and Edward C. Norton. 2003. “Choosing between and Interpreting the Heckit and Two-Part Models for Corner Solutions.” *Health Services and Outcomes Research Methodology* 4 (1). Kluwer Academic Publishers: 5–18. doi:10.1023/A:1025827426320.

Durtschi, Cindy, William Hillison, and Carl Pacini. 2004. “The Effective Use of Benford ’ s Law to Assist in Detecting Fraud in Accounting Data.” *Journal of Forensic Accounting* 99 (99): 17–34. doi:DOI:

Dye, Ronald A. 1986. “Proprietary and Nonproprietary Disclosures.” *The Journal of Business* 59 (2). The University of Chicago Press: 331. doi:10.1086/296331.

Ebrahim, Alnoor. 2003. “Accountability in Practice: Mechanisms for NGOs.” *World Development* 31 (5). Pergamon: 813–29. doi:10.1016/S0305-750X(03)00014-7.

Epstein, Marc J., and Kristi Yuthas. 2014. “Measuring and Improving Social Impacts.” In *Measuring and Improving Social Impacts: A Guide for Nonprofits, Companies and Impact Investors*, 44:0–9. Routledge. doi:10.4324/9781351276245.

Fang, Hanming, and Qing Gong. 2017. “Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked.” In *American Economic Review*, 107:562–91. doi:10.1257/aer.20160349.

Fruttero, Anna, and Varun Gauri. 2005. “The Strategic Choices of NGOs: Location Decisions in Rural Bangladesh1.” *Journal of Development Studies* 41 (5): 759–87. doi:10.1080/00220380500145289.

Fudenberg, Drew, and Jean Tirole. 1995. “A Theory of Income and Dividend Smoothing Based on Incumbency Rents.” *Journal of Political Economy* 103 (1). The University of Chicago Press: 75–93. doi:10.1086/261976.

Gneezy, Uri, Bettina Rockenbach, and Marta Serra-Garcia. 2013. “Measuring Lying

Aversion." *Journal of Economic Behavior and Organization* 93 (September). North-Holland: 293–300. doi:10.1016/j.jebo.2013.03.025.

Hatte, Sophie, and Pamina Koenig. 2018. "The Geography of NGO Activism against Multinational Corporations." *The World Bank Economic Review*. Accessed September 20. doi:10.1093/wber/lhy007.

Healy, Paul M. 1985. "The Effect of Bonus Schemes on Accounting Decisions." *Journal of Accounting and Economics* 7 (1–3). North-Holland: 85–107. doi:10.1016/0165-4101(85)90029-1.

Herzlinger, R E. 1996. "Can Public Trust in Nonprofits and Governments Be Restored?" *Harvard Business Review* 74 (2): 97–107. http://www.ncbi.nlm.nih.gov/pubmed/10156469.

Hill, T.P. 1995. "A Statistical Derivation of the Significant-Digit Law." *Statistical Science*. doi:10.2307/2246134.

Hofmann, Mary Ann, and Dwayne McSwain. 2013. "Financial Disclosure Management in the Nonprofit Sector: A Framework for Past and Future Research." *Journal of Accounting Literature* 32 (1). Elsevier: 61–87. doi:10.1016/j.acclit.2013.10.003.

Jensen, Michael C., and William H. Meckling. 1979. "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure." In , 163–231. Springer, Dordrecht. doi:10.1007/978-94-009-9257-3_8.

Judge, George, and Laura Schechter. 2009. "Detecting Problems in Survey Data Using Benford's Law." *Journal of Human Resources* 44 (1): 1–24. doi:10.1353/jhr.2009.0010.

Keating, Elizabeth K., and Peter Frumkin. 2003. "Nonprofit Reengineering Toward Accountability : A Financial More Reliable Foundation for Regulation." *Public Administration Review* 63 (1). Wiley and American Society for Public Administration: 3–15. doi:10.2307/977517.

Kothari, S. P., Susan Shu, and Peter D. Wysocki. 2009. “Do Managers Withhold Bad News.” *Journal of Accounting Research* 47 (1). Wiley/Blackwell (10.1111): 241–76. doi:10.1111/j.1475-679X.2008.00318.x.

Madden, David. 2008. “Sample Selection versus Two-Part Models Revisited: The Case of Female Smoking and Drinking.” *Journal of Health Economics* 27 (2). North-Holland: 300–307. doi:10.1016/j.jhealeco.2007.07.001.

Mansuri, Ghazala, and Vijayendra Rao. 2012. *Localizing Development*. The World Bank. doi:10.1596/978-0-8213-8256-1.

McCormick, Barry. 1990. “(Search Readings) - A Theory of Signalling During Job Search, Employment Efficiency, and ‘Stigmatised’ Jobs.” *The Review of Economic Studies* 57 (2): 15. doi:10.2307/2297383.

McGuire, Sean T., Thomas C. Omer, and Nathan Y. Sharp. 2012. “The Impact of Religion on Financial Reporting Irregularities.” *The Accounting Review* 87 (2). American Accounting Assocation: 645–73. doi:10.2308/accr-10206.

Michalski, Tomasz, and Gilles Stoltz. 2013. “Do Countries Falsify Economic Data Strategically? Some Evidence That They Might.” *Review of Economics and Statistics* 95 (2): 591–616. doi:10.1162/REST_a_00274.

Miller, Steven J. 2015. *Benford’s Law : Theory and Applications*. Princeton University Press.

Morrow, John. 2014. “Benford’s Law, Families of Distributions and a Test Basis.” *CEP Discussion Paper* No 1291: 1–26. http://cep.lse.ac.uk/pubs/download/dp1291.pdf.

Nigrini, Mark J. 2012. *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations*. *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations*. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9781118386798.

Norton, Edward C., William H. Dow, and Young Kyung Do. 2008. “Specification Tests for the Sample Selection and Two-Part Models." *Health Services and Outcomes Research Methodology* 8 (4). Springer US: 201–8. doi:10.1007/s10742-008-0037-8.

Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2). The University of Chicago Press: 200–249. doi:10.1086/517935.

Olken, Benjamin A., and Rohini Pande. 2012. "Corruption in Developing Countries." *Annual Review of Economics* 4 (1). Annual Reviews: 479–509. doi:10.1146/annurev-economics-080511-110917.

Powell, James L. 1984. "Least Absolute Deviations Estimation for the Censored Regression Model." *Journal of Econometrics* 25 (3). North-Holland: 303–25. doi:10.1016/0304-4076(84)90004-6.

Robinson, P. M. 1988. "Root-N-Consistent Semiparametric Regression." *Econometrica* 56 (4). The Econometric Society: 931. doi:10.2307/1912705.

Schulter, Günter, Erich Mittenecker, and Ilona Papousek. 2010. "A Computer Program for Testing and Analyzing Random Generation Behavior in Normal and Clinical Samples: The Mittenecker Pointing Test." *Behavior Research Methods* 42 (1). Springer-Verlag: 333–41. doi:10.3758/BRM.42.1.333.

Silva, João M. C. Santos, Silvana Tenreyro, and Frank Windmeijer. 2015. "Testing Competing Models for Non-Negative Data with Many Zeros." *Journal of Econometric Methods* 4 (1). De Gruyter: 29–46. doi:10.1515/jem-2013-0005.

Verrecchia, Robert E. 2001. "Essays on Disclosure." *Journal of Accounting and Economics* 32 (1–3): 97–180. doi:10.1016/S0165-4101(01)00025-8.

Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57 (2). The Econometric Society: 307. doi:10.2307/1912557.

Wooldridge, Jeffrey M. 2010. "Econometric Analysis of Cross Section and Panel Data." *MIT Press: Cambridge, MA, USA*. MIT Press, 1064. doi:10.1515/humr.2003.021.

Zitzewitz, Eric. 2012. "Forensic Economics." *Journal of Economic Literature* 50 (3): 731–69. doi:10.1257/jel.50.3.731.

Table 1. Descriptive statistics of Ugandan NGOs

VARIABLES	*Full*		*Incomplete*	*Complete*	
	Mean	S.D	Mean	Deviate	Conform
Measure of accuracy	0.71	0.423	0	0.942	0.964
Performance	4.202	1.037	4.407	3.842	4.224
Reports requested per year	1.288	2.639	0.815	2.105	1.241
% Professional degrees	0.350	0.304	0.282	0.407	0.363
Government as a hindrance	0.413	0.495	0.556	0.368	0.362
Registered as company	0.606	0.491	0.630	0.526	0.621
Received grant	0.673	0.471	0.481	0.579	0.793
Lack of skilled staff	0.529	0.502	0.481	0.579	0.534
Lack of funding	0.731	0.446	0.778	0.632	0.741
Years working in government	5.875	0.495	5.963	7.842	5.190
Tenure length	6.481	8.286	5.667	7.605	6.492
Clerical staff	3.894	4.702	2.111	4.105	4.655
NGO Age of existence	10.442	6.399	11.074	11.000	9.966
Religious Affiliation	0.356	9.700	0.185	0.526	0.379
Board	0.894	0.481	0.926	0.789	0.914
Member involvement	0.404	0.309	0.333	0.474	0.414
% of group aged > 55	0.094	0.138	0.115	0.079	0.089
% of group with connection to NGO	0.346	0.338	0.398	0.283	0.343
Observation	104		27	19	58

Notes: Statistics are means unless otherwise stated. Binary variables take the value of 1 if yes and 0 otherwise. *Incomplete* represents 27 NGOs who only provide either revenues or expenses related information or none. *Complete* represents 77 NGOs who provide all revenue and expenses related financial figures as requested. Categorization of conformity is based on the two-sample Kolmogorov – Smirnov test between the observed distribution and the theoretical distribution (see Appendix B for details). *Source:* Authors' analysis based on the 2002 Ugandan NGO survey data.

Table 2. Estimations of the selection and outcome equations

VARIABLES	Selection equation $C_i(.)$	Outcome equation: R_i Heckit	Cragg's
Performance (community satisfaction)	51.851	3.384***	3.459**
	(165.804)	(1.282)	(1.648)
Board	-849.361**	4.272	3.750
	(406.894)	(4.430)	(4.642)
Registered as company	136.415	9.316***	9.419***
	(370.485)	(2.701)	(2.675)
Received grant (Yes = 1)	808.361***	-0.740	-0.092
	(226.026)	(3.869)	(3.218)
NGO Age	139.149***	1.093	1.181***
	(51.613)	(0.681)	(0.381)
NGO Age^2	-4.573***	-0.033*	-0.035***
	(1.134)	(0.020)	(0.010)
Tenure length	-90.388*	-0.075	-0.112
	(51.517)	(0.336)	(0.186)
% Professional degrees	1,287.665**	-1.015	0.197
	(598.840)	(6.155)	(2.903)
Clerical staff (log)	75.902*	0.019	0.036
	(41.018)	(0.205)	(0.143)
Reports requested per year	48.942	-0.923*	-0.846**
	(78.927)	(0.557)	(0.404)
Religious affiliation	1,574.535***	-4.757	-4.062
	(334.990)	(3.818)	(2.540)
Lack of skilled staff	231.249	-1.997	-1.732
	(336.440)	(2.794)	(2.644)
Lack of funding	-503.696	2.564	2.351
	(369.878)	(2.937)	(2.385)
Government as a hindrance	-739.811**	-2.261	-2.760
	(298.481)	(3.249)	(3.329)
Years working in government	2.964	-0.462**	-0.444*
	(16.210)	(0.174)	(0.231)
Member involvement	762.451**		
	(367.103)		
Number of non-zeros		0.543***	0.549***
		(0.103)	(0.076)
Lambda/sigma		-2.522	10.357***
		(8.924)	(0.647)
Constant	-221.743	912.200***	909.986***
	(733.631)	(10.870)	(9.545)
Observations	104	104	104

Note. *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (14) in parentheses. Selection equation estimation uses logit. Coefficients in the selection equation are multiplied by 1000 for easier interpretation. Jaque-Bera test statistics of residual normality in the outcome equation: Pr(Skewness) = 0.142, Pr(Kurtosis) = 0.3699, Joint-test chi-square statistic (p-value) = 3.03 (0.22). Breusch-Pagan heteroscedasticity statistics (p-value) = 3.91 (0.05). Variable *Performance* is the evaluation from the community focus group. Section 5.2 provides a description of the controls.

Table 3. Model selection tests

The HPC test for the preferred model (Silva, Tenreyro, and Windmeijer 2015)				
Null Hypothesis (Ho)	Heckman model is valid		Cragg model is valid	
t-statistics	3.054***		-3.064	
Probability > t (p-values)	0.001		0.999	
Vuong's (1989) test for non-nested models				
Null Hypothesis (Ho)	The respective distances to the unknown "true" model are equal			
Alternative Hypothesis (H_1)	Cragg's specification is closer			
Ln Ration (s.e) [p-value]	6.753 (0.394) [0.000]			
Observations	104	104	104	104

Note. Both tests unanimously indicate the double-hurdle censoring mechanism fits the data better. See Online Appendix B for the details of the tests.

Table 4. IV estimations of the selection and outcome equations

Panel A. Estimates of the selection and outcome questions

VARIABLES	Selection equation $C_i(.)$	Outcome equation: R_i IV-Heckit	Outcome equation: R_i IV-Cragg
Performance (community satisfaction)	271.502	10.92**	11.289**
	(377.653)	(5 .114)	(5.154)
Anderson-Rubin coverage-corrected confidence interval (p-value)		[0.608, 60.06] (0.040)	
Lambda/sigma		15.209	10.089***
		(11.054)	(0.634)
Observations	100	100	100

Panel B. Diagnostic test for IV first stage estimation

INSTRUMENTAL VARIABLES	*Dependent variable:* Performance	
% of group aged > 55	-1.635 ***	(0.585)
% of group with connection to the NGO	0.792 **	(0.406)
Sanderson-Windmeijer F test of excluded instruments: (Prob > F)	4.49***	(0.016)
Cragg-Donald Wald F-statistics (weak identification)		4.485
Kleibergen-Paap rk LM statistic (under-identification) (p-value)	8.920 **	(0.012)
Hansen J statistics (overidentification) (p-value)	0.596	(0.440)
Anderson-Rubin Wald weak-instrument-robust test: (Prob>F)	2.48	(0.095)

Note. *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (14) in parentheses unless stated otherwise (bootstrapped standard errors with 100 replications are largely identical). The coefficient θ_2 of the predicted value of the suspected endogenous variable ($\hat{v}$) = -9.212 * (5.160), indicating the presence of endogeneity. Anderson-Rubin coverage-corrected confidence interval and p-value are based on approximations. The null hypothesis of Anderson-Rubin Wald weak-instrument-robust inference is that coefficients of instruments are insignificant in the structural equation and the orthogonality conditions are valid. The test is robust to potential weak instrumentation. The null hypothesis of Sanderson-Windmeijer's (2016) F-test statistics is that the instruments can be excluded from the first-stage estimation. The null hypothesis of Cragg-Donald (1993) Wald test statistics is that the instruments are weakly identified when compared against Stock-Yogo weak ID test critical value with the maximal LIML size of tolerance bias at 20% (4.42). Rejection of these null hypotheses suggests the absence of a weak instrumentation problem. The over-identification and under-identification tests hypothesize if the instrumentation is identified and under-identified.

Table A1. Probability predicted by Benford's Law for the leading digits

d	1	2	3	4	5	6	7	8	9
$P(d)$	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Table A2. Estimations of outcome equations for alternative measures of reporting accuracy

VARIABLES	*d*-statistics		KS statistics		*Conform*
	Heckit	Cragg	Heckit	Cragg	Heckit
Performance	1.020***	1.010**	12.534**	11.606*	92.360*
	(0.334)	(0.497)	(5.601)	(5.972)	(47.541)
Board	0.910	0.980	33.099*	39.582	146.372
	(1.153)	(1.293)	(19.335)	(25.962)	(164.118)
Registered as company	2.699***	2.685***	20.812*	19.531*	131.501
	(0.703)	(0.723)	(11.741)	(11.847)	(99.671)
Received grant (Yes = 1)	-0.301	-0.388	28.745*	20.693	287.253**
	(1.008)	(0.816)	(16.663)	(15.351)	(141.492)
Age	0.353**	0.341***	4.481	3.391**	39.959
	(0.177)	(0.109)	(2.949)	(1.358)	(25.035)
Age^2	-0.011**	-0.010***	-0.166*	-0.131***	-1.414*
	(0.005)	(0.003)	(0.086)	(0.032)	(0.729)
Tenure length	-0.012	-0.007	-1.842	-1.388	-18.429
	(0.087)	(0.054)	(1.475)	(1.183)	(12.520)
% Professional degrees	-0.181	-0.343	-3.847	-18.893	145.409
	(1.604)	(0.807)	(26.526)	(16.889)	(225.240)
Clerical staff	0.012	0.009	-0.209	-0.419	1.847
	(0.053)	(0.044)	(0.901)	(0.732)	(7.650)
Reports requested per an.	-0.150	-0.161	-3.960	-4.917**	-20.833
	(0.145)	(0.100)	(2.432)	(1.912)	(20.643)
Religious Affiliation	-0.704	-0.797	3.322	-5.320	-65.545
	(0.994)	(0.741)	(16.698)	(11.529)	(141.723)
Lack of skilled staff	-0.322	-0.358	4.244	0.950	-44.901
	(0.728)	(0.680)	(12.112)	(13.060)	(102.829)
Lack of funding	0.589	0.618	5.824	8.473	54.232
	(0.764)	(0.607)	(12.815)	(9.443)	(108.775)
Government as a hindrance	-0.743	-0.676	-29.967**	-23.760	-115.436
	(0.846)	(0.725)	(14.059)	(19.908)	(119.363)
Years working in government	-0.113**	-0.110*	-1.159	-0.911	-11.653*
	(0.045)	(0.061)	(0.759)	(0.873)	(6.440)
Number of non-zeros	0.049*	0.048***	1.567***	1.497***	5.724
	(0.027)	(0.016)	(0.429)	(0.299)	(3.646)
Constant	30.319***	30.67***	701.5***	729.0***	-31.146
	(2.833)	(2.177)	(46.806)	(40.521)	39.745
Observations	104	104	104	104	104

Note. *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (14) in parentheses. The selection equation is identical as in Table 1. We omit the results the double-hurdle model for *Conform* due to the binary nature of the dependent variable.

Table A3. A balance test for correlations between NGO characteristic and instruments

VARIABLES	IV-connection	IV-senior
Reports requested per year	0.001	-0.005
	(-1.161)	(-0.452)
% Professional degree	21.123**	-0.015
	(-9.38)	(-3.724)
Registered as company	-9.88*	0.000
	(-5.852)	(-2.353)
Received grant	-0.093	-6.47***
	(-6.112)	(-2.33)
Lack of skilled staff	-0.110	0.001
	(-6.682)	(-2.633)
Lack of funding	0.014	-0.031
	(-5.834)	(-2.244)
Government as a hindrance	-0.050	0.002
	(-5.944)	(-2.35)
Years working in government	-0.001	0.002
	(-0.385)	(-0.148)
Tenure	-0.006	-0.002
	(-0.604)	(-0.236)
Clerical staff	0.000	0.000
	(-0.011)	(-4.18E-03)
Age	-0.003	0.000
	(-0.00256)	(-0.001)
Religious Affiliation	-12.043*	-0.012
	(-6.161)	(-0.0243)
Board	0.013	0.046
	(-9.363)	(-3.65)
Member involvement	0.072	0.001
	(-5.89)	(-2.314)

Note. Bootstrapped standard errors in brackets. Coefficients reported from separate linear regressions, where each characteristic is regressed on the respective instrument by OLS. There are generally no insignificant associations between NGO characteristics and each instrument. We show here that the instruments do not exhibit any strong significant correlations with observable characteristics of the NGO working in the community. All estimates are multiplied with 100 for easier presentation.

Table A4. A sample of Financial Information Requested in the 2002 survey.

All amounts in thousands of shillings		2001 or preceding fiscal year	2000 or preceding fiscal year
REVENUES			
A. Recurrent revenue		,000	,000
Grant from: International NGO	Disbursed	,000	,000
	Authorized	,000	,000
Ugandan NGO	Disbursed	,000	,000
	Authorized	,000	,000
National government	Disbursed	,000	,000
	Authorized	,000	,000
Local government	Disbursed	,000	,000
	Authorized	,000	,000
UN organization	Disbursed	,000	,000
	Authorized	,000	,000
Membership fees		,000	,000
Fees paid by recipients of services rendered by [NGO]		,000	,000
Income from services rendered to the government		,000	,000
Income from services rendered to another NGO/CBO		,000	,000
Income from services rendered to another NGO/CBO		,000	,000
Profit on special events (e.g., fair, concert, paying dinner)		,000	,000
Voluntary donations from members		,000	,000
Voluntary donations from Non-members		,000	,000
Property income/endowment income		,000	,000
Tax refunds		,000	,000
Other: specify		,000	,000
B. Divestment		,000	,000
Sale of land or buildings		,000	,000
Sale of vehicles		,000	,000
Sale of equipment or machinery (including computers)		,000	,000
Total revenues		,000	,000
EXPENDITURES			
A. Recurrent expenditures			
Program costs (what goes to the community: drugs, school books, etc)		,000	,000
Wages/salaries/honorarium or full package		,000	,000
Housing allowances		,000	,000

Transport allowances	,000	,000
Subsistence allowances and per diems to [NGO] staff and volunteers	,000	,000
Subsistence allowances and per diems to [NGO] beneficiaries	,000	,000
Utilities (electricity, water, etc.)	,000	,000
Petrol/fuel	,000	,000
Rent for land and buildings	,000	,000
Payment to NGOs/CBOs for services rendered	,000	,000
Payment to someone else for services rendered	,000	,000
Grants/voluntary donation/contribution to another NGO/CBO/Church	,000	,000
Interest on debt and financial charges (including leasing)	,000	,000
Bribes	,000	,000
Other costs	,000	,000
B. Investments	,000	,000
Land and buildings	,000	,000
Vehicles	,000	,000
Equipment and machinery (including computers)	,000	,000
Total annual expenditures	**,000**	**,000**
BORROWING SITUATION		
Loans/borrowed money from:	,000	,000
Ugandan/International NGO Borrowed	,000	,000
Remaining	,000	,000
Local/national Government Borrowed	,000	,000
Remaining	,000	,000
UN agency/Bilateral donor Borrowed	,000	,000
Remaining	,000	,000
Bank/financial institution Borrowed	,000	,000
Remaining	,000	,000
OTHER FINANCIAL INFORMATION		
Amount of Investment financed by hire-purchase	,000	,000
Values of equipment	,000	,000
Values of inventories	,000	,000
Values of vehicles	,000	,000

Note: 1 US dollar = 1806.15 Ugandan Shillings in July 31 2002. All of these financial items are used to conduct our indices using Benford's Law. There are 60 pieces of financial information (financial items) requested for each NGO in each year. Ideally, we would be able to collect 120 financial items over two years to construct measures of information accuracy using Benford's Law.

Figure Legends

There is no legend for Figure 1 and Figure 2.

Figure Legend for Figure 3

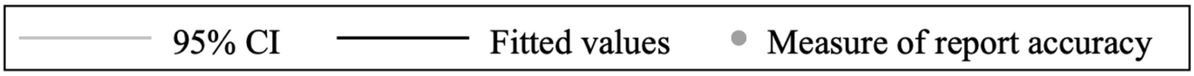


Figure Legend for Figure A1

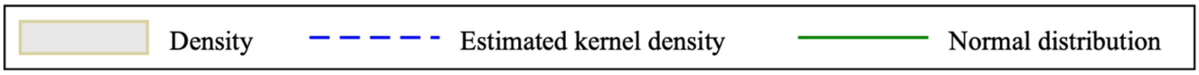
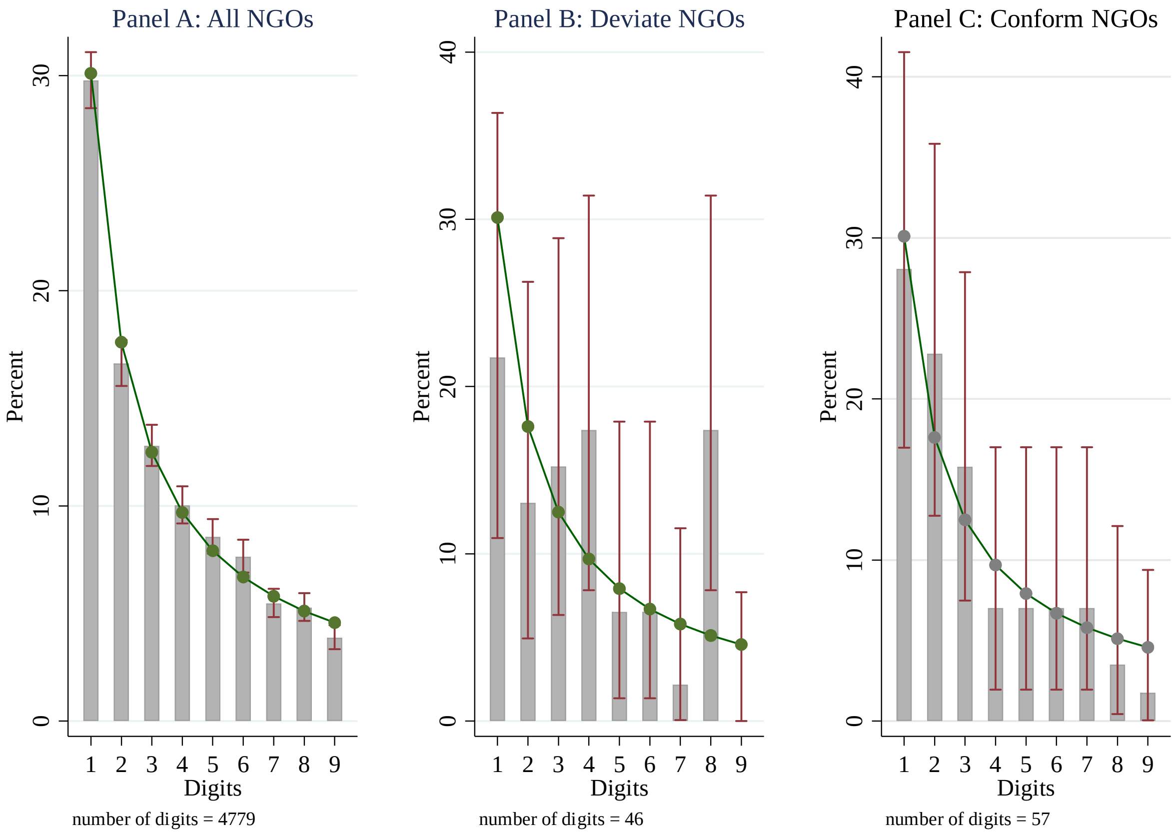
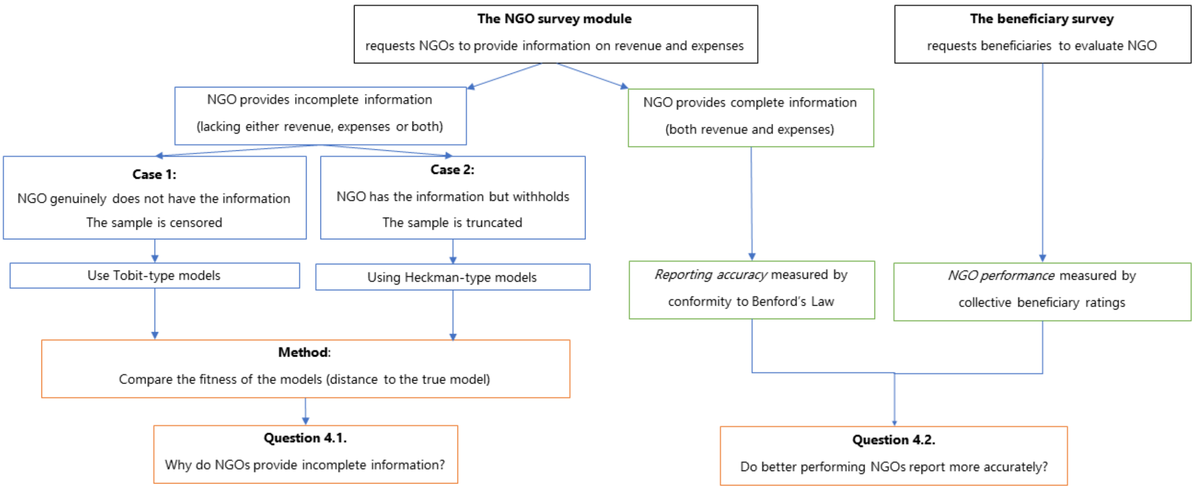


Figure 1. The Ugandan NGO financial data and conformity to Benford's Law



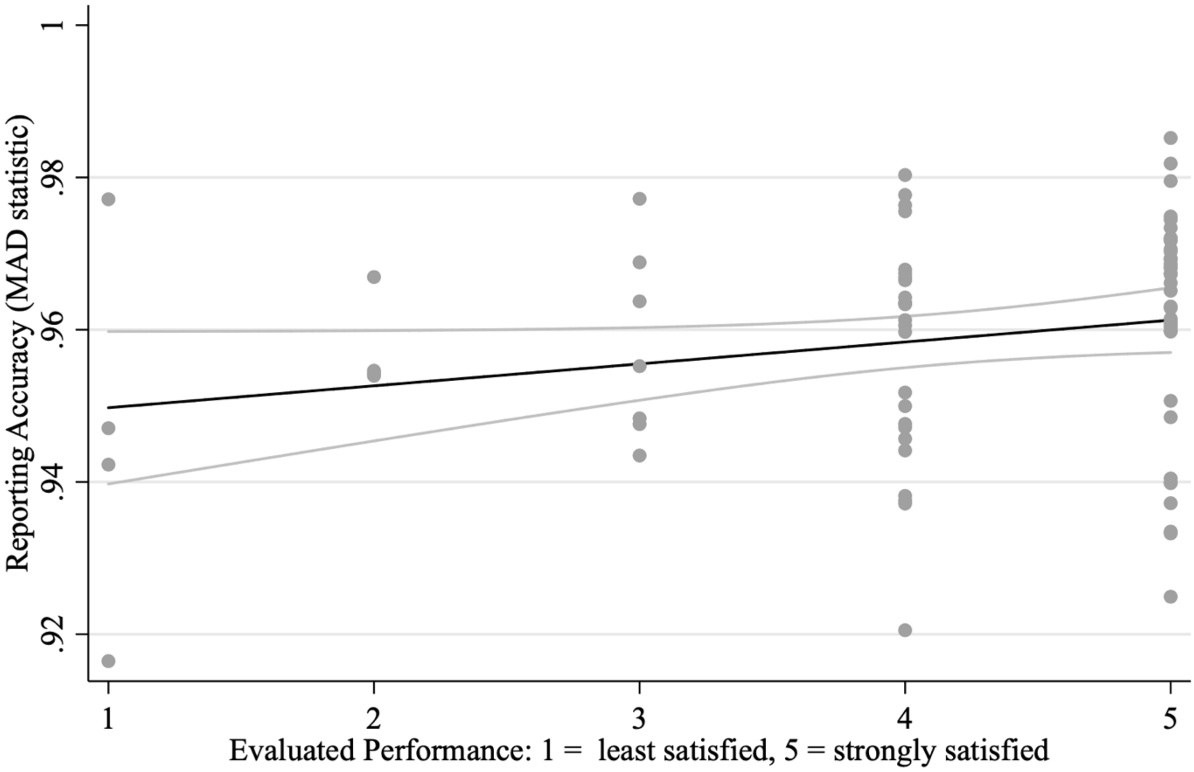
Note. Lines represent the theoretical frequency of digits 1 to 9 appearing as the first digits according to Benford's Law. Bars represent the observed distributions in the three samples. Capped spikes represent confidence intervals at the 10% level of significance. Panel A presents the distributions for all the numbers in all financial accounts provided by the NGOs. Panel B is a representative NGO (25% of the sample) whose requested financial accounts fail the hypothesis test of conformity to the law using the two-sample Kolmogorov – Smirnov test. Panel C is a representative NGO (75% of the sample) where we fail to reject the hypothesis test of conformity. *Source.* Authors' analysis based on the 2002 Ugandan NGO survey data.

Figure 2. Flowchart illustrates steps of our analysis



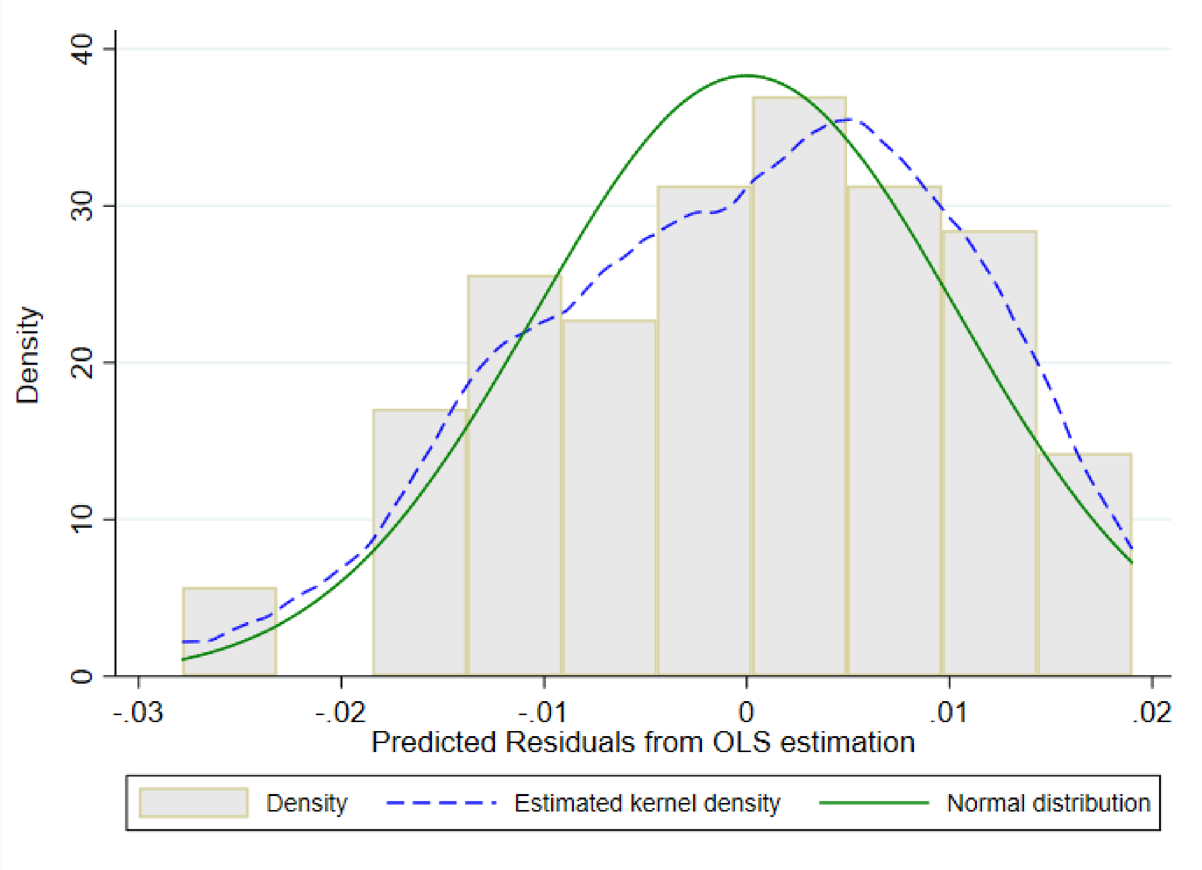
Note. A possible Case 3 is that incomplete information is due to the lack of authorization or the enumerator effect. We address this concern with the survey design and checking with interviewer fixed effect in Footnote 11.

Figure 3. Descriptive relation between the degree of report accuracy and NGO performance



Note. OLS regression using the full sample from the matched 2002 Ugandan NGO survey and corresponding community data. Measure of report accuracy is the MAD calculated using Benford's Law for NGOs with complete disclosure, Evaluated Performance is from the community focus groups.

Figure A1. Density plot of predicted residuals from OLS estimation of the main equation



Note. Estimated kernel density is also plotted. A clear resemblance between the kernel density and the normal distribution suggests the normality condition in residual terms is graphically valid.

Supplementary Online Appendix for Online Publication

Better performing NGOs do report more accurately: Evidence from investigating Ugandan NGO financial accounts

by

Ronelle Burger (Stellenbosch University)

Canh Thien Dang (University of Nottingham)

Trudy Owens (University of Nottingham)

****

A. Conceptualizing the mechanisms of information disclosure and accuracy

We model the decisions made by an (imperfectly) altruistic NGO *i* on whether to provide all requested information and then their report's accuracy level to a third party. The model complements the conceptual framework in Section 4. There are two main results. First, an NGO may provide incomplete information because: (i) the NGO strategically withholds information even when all information is available; and (ii) the NGO chooses *ex ante* to record only some necessary information, hence unrecorded data is simply due to the data having never been collected. We further show that the higher the *ex-ante* cost of gathering information, the lower the accuracy of sequential reports submitted by the organization. This proposition is consistent with the finding that costly accounting procedures could harm the monitoring process as the organization now faces a higher trade-off between upward accountability and downward accountability. Once the organization's preference is to mainly focus on the utility of their beneficiaries rather than accountability (either due to their antagonistic attitude or the donors being lenient toward accountability), incomplete disclosure remains an option.

The NGO decides by solving the following maximization problem:[19]

$$\max_{P_i, R_i} U_i(P_i, R_i) \ \ s.t\ E_i = P_i + R_i + \tau C(.) \tag{OA.1}$$

where U_i is a (possibly individual-specific) well-behaved, continuous and twice differentiable utility function on its domains $P_i, R_i \in \mathbb{R}_{\geq 0}$ indicating the outcome of their altruistic projects (e.g. feedback scores from their beneficiary community) and the degree of report accuracy (for example, among 100 reported items, how many are recorded accurately), respectively.[20] The term $E_i \in \mathbb{R}_{>0}$ designates the positive, fixed resources of the NGO (including non-monetary effort of the manager). The NGO decides to spend on delivering altruistic projects and determining the degree of report accuracy. As discussed above, we implicitly assume that the NGO can choose the extent of its reporting accuracy. This is not unreasonable because an NGO can either exert more resources to record detailed transactions and avoid human errors (increased diligence) or simply have increased integrity. For tractability, we simply assume that the outcome of their altruistic projects and reporting with a degree of $R_i > 0$ require the same numeric amount R_i of resources. If the NGO spends $R_i = 0$, we have $C(.) = 0$ or *incomplete disclosure*. Reporting with a degree of R_i further incurs a fixed cost of information acquisition $\tau C(.)$. The parameter $\tau > 0$ reflects the increasing cost of information gathering (e.g. hiring at least a clerk to manage book-keeping). We denote $C_i(.)$ as an indicator function such that $C_i(.) = 1$ if NGO *i* provides a complete set of the requested

[19] The illustration is inspired by the selection mechanism in (Bettin, Lucchetti, and Zazzaro 2012) on remittances.

[20] Another way to interpret the degree of report accuracy is the probability of the report being found accurate by an objective test. We capture this interpretation by using the p-values of Chi-square and Kolmogorov–Smirnov tests of whether the reports follow Benford's Law as measures of report accuracy.

information (*complete disclosure*), and $C_i(.) = 0$ if either expenditure-related or revenue-related information or both are missing (*incomplete disclosure*). In the case of incomplete disclosure, we treat NGO i as exerting no resources in reporting or $R_i = 0$ if $C_i(.) = 0$. Otherwise, $C_i(.) = 1$ if $R_i > 0$.[21] The parameter $\tau > 0$ represents the fixed cost of disclosing full information, which includes either costs of information acquisition or ex post costs discussed above.

To reflect the altruistic aspect, we further assume:

- $U_i(0, R_i) = 0$ and $U_i(P_i, 0) > 0$ and $U_i(P_i, .)$ is a strictly quasi-concave function of P_i. This set ensures that the altruistic NGO derives no utility from diverting all the given resources away from delivering core projects and they always gain from completely focusing on altruistic activities. The last assumption is to imply the NGO's preference on the consumption set of performance measures (such as beneficiary's feedback) is convex.[22]
- $U_P^i = \frac{\partial U_i}{\partial P_i} > 0$, $U_{PP}^i > 0$, $U_R^i = \frac{\partial U_i}{\partial R_i} \geq 0$, $U_{RR}^i > 0$. This set ensures that increased performance measure and reporting accuracy provide increasing marginal utility. Note that by $U_R^i \geq 0$, we also implicitly assume that some NGOs may gain zero additional utility from increased accuracy.

[21] We rule out cheap talk by implicitly assuming that a full disclosure carries some extent of true information.

[22] Formally, let $x, y \in X$ denote two performance measure values of a set X and $y \succcurlyeq x$, then for every $t \in (0,1)$: $ty + (1-t)x \succcurlyeq x$. That is, if y is preferred over x, then any mix of the two is still preferred than x.

Proposition 1. If $\tau > 0$, there always exists a unique solution $(P_i, R_i) \in \mathbb{R}_{\geq 0}$ to (1).[23] Furthermore, the NGO may choose *incomplete disclosure*, $C_i(.) = 0$, in two situations that correspond to the explanations in Section 4.1:

i. Either when $U_R^i = 0$ or when the non-zero optimal report accuracy the NGO plans to have is feasible but so low such that the utility of incomplete disclosure outweighs the potential optimal utility. Formally, there exists a reservation level of accuracy $\underline{R} > 0$ that for all $0 < r_i \leq \underline{R}$: $U_i(p_i, r_i) < U_i(E_i, 0)$. Incomplete disclosure is preferred since $R_i^* \leq \underline{R}$.

ii. If the fixed cost of disclosure τ is sufficiently high that a non-negative accuracy is not feasible.

Proof. If $\mathrm{U_R^i} = 0$, the utility function is constructed such that the NGO's preference does not attach any additional utility to either increased accuracy or complete disclosure. The problem has a unique solution $(E_i, 0)$.[24] In this case, the construction of U_R^i governs both mechanisms underlying disclosure and accuracy decisions.

If $\mathrm{U_R^i} > 0$, we can solve the two following auxiliary problems by Karush-Kuhn-Tucker conditions:

$$\max_{P_i} U_i(P_i, E_i - \tau - P_i) \quad s.t \qquad E_i - \tau \geq P_i \geq 0 \tag{OA.2}$$

$$\max_{P_i} U_i(P_i, E_i - P_i) \qquad s.t \qquad E_i \geq P_i \geq 0 \tag{OA.3}$$

[23] The proposition does not hold if $\tau < 0$. For example, let $E_i = 1$ and $\tau = -1$ and $U_i(P_i, R_i) = 4P_i^2 + R_i^2 + P_i + R_i$. As the function defining the constraint set is discontinuous at $(2,0)$, the maximisation problem has no solution.

[24] Another situation by construction is when $U_i(.,.)$ is convex on its domains, or $U_{PP}U_{RR} - U_{PR}^2 \geq 0$. In that case, $(E_i, 0)$ maximises the original problem as $U_i(E_i - \tau, 0) < U_i(E_i, 0)$ by $U_P > 0$.

Let x_1 and x_2 be the respective solutions of (OA.2) and (OA.3), u_1 and u_2 be the respective values of the maximized utility. Denote $G(P_i) = E_i - \tau C(.) - P_i$ the constraint set of the main problem. As $\tau > 0$, $\{P_i: G(P_i) \geq 0, x \geq 0\}$ is non-empty and compact. According to Wierstrass Theorem, since $U_i(P_i, .)$ is continuous on P_i, there exist $P_1 = x_1$ and $P_2 = x_2$ that solve (OA.2) and (OA.3 respectively.

We now show x_1 and x_2 are unique. For x_1, suppose there exists two maxima $x_1 \neq x_1'$ $s.t$ $U_i(x_1, .) = U_i(x_2, .), G(x_1) \geq 0, G(x_1') \geq 0$. For $t \in (0,1)$:

$$G(tx_1 + (1-t)x_1') = tG(x_1) + (1-t)G(x_1') \geq 0 \tag{OA.4}$$

Thus, $\mathrm{tx}_1 + (1-\mathrm{t})\mathrm{x}_1'$ is a feasible point. Since $\mathrm{U}_\mathrm{R}^\mathrm{i} > 0$ and $\mathrm{U}_\mathrm{i}(\mathrm{P}_\mathrm{i}, .)$ is strictly quasi-concave on P_i, we have that:

$$U_i(tx_1 + (1-t)x_1', .) > \min\{U_i(x_1, .), U_i(x_2, .)\} = u_1 \tag{OA.5}$$

This is a contradiction as u_1 is assumed the maximised value, or x_1 is unique. A similar rationale applies for x_2.

To specify the solutions for the original problem we compare u_1 with $U_i(E_i, 0)$, that is when the NGO exerts no resources on reporting. If $u_1 > U_i(E_i, 0)$ and $E_i - \tau > x_1$, x_1 solves the original solution. In other words, the NGO discloses and chooses some non-negative level of inaccuracy at optimum (since $R_1 = E_i - \tau - x_1 > 0$). Otherwise, we have two situations that lead to a solution of incomplete disclosure.

First, if $\mathrm{u}_1 = \max\limits_{E_i - \tau > P_i \geq 0} U_i(P_i, R_i) < U_i(E_i, 0)$, the solution to the main problem must be either $(P_i, 0)$ or $(x_2, 0)$ depending on which utility between $U_i(E_i, 0)$ and u_2 is larger. Notice that since $\max\limits_{E_i - \tau > P_i \geq 0} U_i(P_i, R_i) < U_i(E_i, 0)$, $\tau > 0$ and $U_R^i > 0$, we have $\lim\limits_{R_i \to 0} U_i(P_i, R_i) < U_i(E_i, 0)$. Thus, there exists a reservation level of report accuracy $\underline{R} > 0$ such that for all $0 < r_i \leq \underline{R}$: $U_i(p_i, r_i) < U_i(E_i, 0)$. Combining with $u_1 = \max\limits_{E_i - \tau > P_i \geq 0} U_i(P_i, R_i) < U_i(E_i, 0)$, we interpret this as the optimal report accuracy that the NGO plans to have is so low that the utility of incomplete disclosure outweighs the potential optimal utility. Either way, *incomplete disclosure* is the solution.

Second, if $u_1 > U_i(E_i, 0)$ but $E_i - \tau < x_1$, hence (x_i, R_i^*) is not feasible. Again, the solution to the main problem must be corner and be either $(P_i, 0)$ or $(x_2, 0)$ depending on the relative value of $U_i(E_i, 0)$ and u_2. The underlying mechanism for the corner solution is however due to the maximised value of u_1 so that a feasible R_i^* is unattainable (in this case, it must be negative). The intuition is that as the fixed cost τ is set too high that $E_i - \tau < x_1$ for NGO *i*, the optimal report accuracy must have been negative for the NGO. The NGO maximizes utility by choosing $\mathrm{R_i} = 0$. *QED*

If we further assume that $\mathrm{U_\tau} < 0$, that is the NGO is worse off when the fixed cost of information gathering increases (e.g. they are left with fewer resources for charitable activities). We have the following Proposition 2.

Proposition 2. If $\mathrm{R_i^*} \in \mathbb{R}_{\geq 0}$ is the optimal reporting accuracy that solves the main problem, then $\frac{\partial \mathrm{R_i^*}}{\partial \tau} \leq 0$. That is, as the fixed cost of disclosure increases, the optimal choice of accuracy decreases.

Proof. We formally show that $\mathrm{R_i^*}$ exists in the proof of Proposition 1. Given the existence, if incomplete disclosure occurs, $\mathrm{R^*} = 0$, the lemma is bounded.

If complete disclosure occurs, consider the main maximization problem with respect to $\mathrm{P_i}$ over $[0, \mathrm{E_i} - \tau_\mathrm{i}]$. Under the (bounded) lattice constraint, rewrite the maximisation in terms $\mathrm{R_i}$: $\max\limits_{0 \leq \mathrm{P_i} \leq \mathrm{E_i} - \tau} \mathrm{U_i}(\mathrm{E_i} - \tau - \mathrm{R_i}, \mathrm{R_i})$.

Since we only have one choice variable, super-modularity is trivial. As the constraint set is a bounded lattice, we will only need to check increasing differences in $(\mathrm{R_i^*}, -\tau = \vartheta)$. Take partial derivatives of $\mathrm{U_i}$, we get $\mathrm{U_{R_i^*\vartheta}} = \frac{\partial \mathrm{U_{R_i}}}{\partial \vartheta} = \frac{\partial \mathrm{U_{R_i}}(\mathrm{E_i} - \tau - \mathrm{R_i}, \mathrm{R_i})}{\partial \vartheta} > 0$ since $\mathrm{U_\tau} < 0$. Topkis's Theorem suggests that $\frac{\partial \mathrm{R_i^*}}{\partial \tau} < 0$. *QED.*

B. Vuong's (1989) non-nested hypothesis test

As both Heckman and Cragg models are non-nested, we use a Vuong (1989) test to compare the difference in their respective Kullback-Leibler information criterion (KLIC) distance from the unknown "true" model that best fits the data. The distance is defined as follows:

$$KLIC \equiv E(L^{True}) - E(L^{*}) \tag{OA.6}$$

where L^{True} is the log of the conditional density of the unknown true model and L^{*} is the log of the conditional density of the model approximating the data. Vuong (1989) suggests that to minimize KLIC is equivalent to maximizing the expected log-likelihood $E(L^{*})$ and derives the following likelihood statistics adapted in our context:

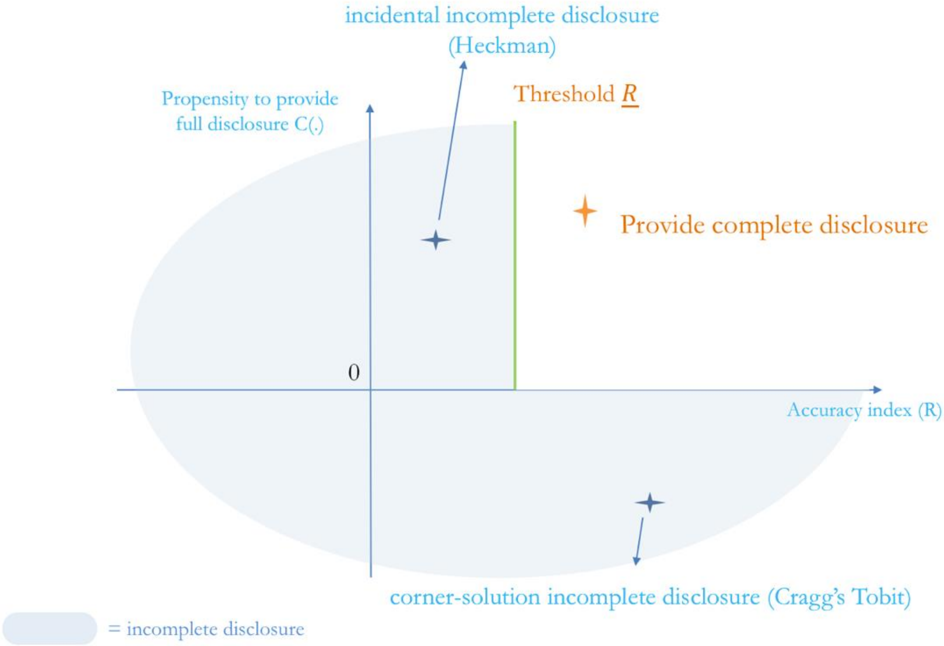
$$z = \frac{LR_n(\hat{\beta}_n, \hat{\theta}_n)}{\widehat{\omega}_n \sqrt{N}} \equiv \frac{\left[L_n^{Cragg}(\hat{\beta}_n) - L_n^{Heckman}(\hat{\theta}_n)\right]}{\widehat{\omega}_n \sqrt{N}} \tag{OA.7}$$

where L_n represents the log-likelihood of the Cragg and Heckman models, $\hat{\beta}_n, \hat{\theta}_n$ are respectively the regressors in the main equation of the two models. $\widehat{\omega}_n^2$ is the estimated variance of the pointwise log-likelihood ratio calculated as:

$$\widehat{\omega}_n^2 \equiv \frac{1}{n}\sum_{1}^{n}\left[\ln\frac{f(\hat{\beta}_n)}{f(\hat{\theta}_n)}\right]^2 - \left[\frac{1}{n}\sum_{1}^{n}\ln\frac{f(\hat{\beta}_n)}{f(\hat{\theta}_n)}\right]^2 \tag{OA.8}$$

where $f(\hat{\beta}_n)$ and $f(\hat{\theta}_n)$ are the individual log-likelihoods of the Cragg and Heckman models. The likelihood statistic z is tested against the standard normal distribution. A positive z suggests that Cragg's model is closer to the unknown true model. Otherwise the Heckman model is preferred. Figure OA1 illustrates the three categories.

Figure OA1. Three types of NGOs regarding information disclosure and report accuracy



Source. Adapted from Bettin et al.'s (2012) example on remittance behaviors.

C. Cragg's and Heckman model with endogenous regressors

Let Z_i be a set of instrument variables, the two equations of interest with the endogenous explanatory variable, x, are written as:

$$R_i^* = \alpha_1 + \gamma x + X_i'\beta_1 + u_i \tag{OA.9}$$

$$C_i(.) = \alpha_2 + +S_i'\beta_2 + v_i \tag{OA.10}$$

Some studies have addressed the simultaneity problems of heterogeneity and selectivity (see Wooldridge, 2010). Blundell and Smith (1994) and Newey (1987) provide the general framework for a control function approach to estimate sample selection and double-hurdle models with endogenous covariates. For some applications, Semykina and Wooldridge (2010) and Schwiebert (2015) develop Heckman selection models with endogenous explanatory variables. Although we acknowledge our small sample size, we adopt Semykia and Wooldridge's (2010) procedure for the IV-Heckman as follows: first estimate a Probit for the selection indicator on instruments $\mathrm{Z_i}$ and other exogenous variables using all observations: $\mathrm{C_i(.) = Z_i'\gamma_1 + S_i'\alpha_1 + u_{1i}}$. Obtain the estimated inverse Mills ratios: $\hat{\lambda}_{i2}$. Second, estimate the adjusted main equation $\mathrm{R_i^* = X_{1i}'\beta_2 + \beta_{IV-Heckman}x + \theta_1\hat{\lambda}_{i2} + v_{i2}}$ by LIML using instruments

($Z_i, \hat{\lambda}_{i2}$) using the selected sample of NGOs who fully disclose. We use LIML instead of 2SLS to improve the efficiency and to avoid potential severe biasedness of 2SLS with weak instruments in small sample size. The standard errors are clustered at district level and bootstrapped with 100 replications.

Other studies address endogeneity issues in Tobit-type models. Smith and Blundell (1986) and Rivers and Vuong (1988) discuss asymptotically efficient two-step maximum likelihood estimators and provide estimation procedures for Tobit and bivariate probit models. Although the procedures are not designed specifically for double-hurdle models, Blundell and Smith (1994) suggest that their approach – discussed in Smith and Blundell (1986) – can be extended to a double-hurdle model by using the appropriate maximum likelihood function specified in Cragg (1971). Another advantage of the Smith-Blundell procedure is it does not require any distributional assumptions for the first stage estimation. For these results, we adopt Blundell-Smith procedure by MLE as our IV-Cragg estimator (Wooldridge 2010 p. 682): first, estimate the reduced form of x on instruments Z_i and other exogenous variables by OLS: $x = Z_i'\gamma_2 + S_i'\alpha_2 + u_{2i}$. Obtain estimated parameters of the OLS residuals of x as: $\hat{v} = x - Z_i'\hat{\gamma}_2 + S_i'\hat{\alpha}_2$. Second, estimate a standard Cragg's double-hurdle model with the main equation as: $R_i^* = X_{1i}'\beta_3 + \beta_{\text{IV-Cragg}}x + \theta_2\hat{v} + v_{2i}$. The estimates are consistent, and the standard errors are clustered at district level and bootstrapped with 100 replications.

The estimates $\beta_{\text{IV-Heckman}}$ and $\beta_{\text{IV-Cragg}}$ from Equation (9) and (12) are the parameters of interest. To test for the presence of endogeneity in x, we use standard t-test on $\hat{\lambda}_{i2}$ and $\hat{v}$. That is, if we reject either of the null hypotheses $H_o: \theta_1 = 0$ or $H_o: \theta_2 = 0$, we also reject the null hypothesis that x can be treated as exogenous in our specification, equivalent to the Hausman test in an ordinary IV linear model. To assess the possibility of weak instruments we report several statistics: (i) the Sanderson-Windmeijer F-test of excluded instruments, computed in the first-stage estimation via OLS; (ii) the Cragg-Donald Wald F statistics against the Stock-Yogo weak ID test critical value of tolerance bias at 10% (15%, 20%) maximal LIML size; (iii) the Kleibergen-Paap rk LM test statistics for H_o: the specification is underidentified; (iv) Anderson-Rubin Wald weak-instrument-robust inference test for coefficients of instruments being insignificant in the structural equation; and (v) Hansen J statistic test for overidentification. We note that these statistics are not technically equivalent for nonlinear models and only provide informative indications (see Sanderson and Windmeijer 2016).

D. Robinson (1988) and Powell (1984) semi-parametric and non-parametric estimators

We estimate two simple nonparametric kernel regressions (local polynomial smooth and lowess smother) in Figure OA2 for a bivariate relationship between performance and reporting accuracy.[25] The nonparametric becomes complex when adding more variables into a kernel regression because it introduces locally sparse noises ("curse of dimensionality").[26] We propose two exercises to rectify. First, we repeat the above nonparametric kernel regressions between the accuracy measure (y) and residuals (u) from an OLS of (y) on all control variables except the evaluated performance (x). The intuition is that the residuals (u) can capture variations in the accuracy measure (y) that are probably due to the excluded variable (x), but not the control variables. A positive relationship between (u) and (y) indicates a positive relationship between the excluded (x) and (y). Both panels in Figure OA3 support the result.

Second, we perform a Robinson (1988) semiparametric estimator for the sample selection model and Powell (1984) censored least absolute deviations (CLAD) for the censored Tobit model (see Online Appendix D). The estimators are robust to heteroscedasticity,

---

[25] The lowess smoother accounts for the values of the evaluated performance variable locating mainly to the right spectrum: only a few NGOs were rated least satisfied.

[26] Das, Newey, and Vella (2003) develop theoretical nonparametric estimators that also allow for endogeneity. Newey and Powell (2003) propose a two-step nonparametric method which avoids the strong exogeneity assumptions. Blundell and Powell (2004) provides a review on nonparametric and semiparametric models dealing with endogeneity.

consistent, and asymptotically normal for a wide class of error distributions. Figure A4 reports Robinson (1988) estimations for scenarios when $Performance_i$ is considered exogenous and endogenous. Using Powell's (1984) estimator, we obtain the estimate for $Performance_i$ of 5.381 with a bootstrapped and clustered standard error of 3.318. The bias-corrected confidence interval is [4.616, 16.517]. The estimation shows a significantly positive relationship between the two variables, assuring us that the results are not sensitive to the distributional and functional-form assumptions made under the Heckman and Cragg's model.

To illustrate Robinson's (1988) estimator we rewrite the selection model as follows:

$$\mathrm{R_i} = \mathrm{m(x_i)} + \mathrm{X_i'}\beta + \lambda(\mathrm{S_i}) + \varepsilon_\mathrm{i} \quad \text{(OA.11)}$$

where $\mathrm{x_i}$ is the evaluated performance variable which enters the equation as the nonparametric component $m(.)$, ruling out the functional dependence with R_i. X_i and S_i are the parametric components of the equation, consisting of other covariates for the outcome and selection equation. $\lambda(S_i)$ is the inverse Mills ratio obtained from the selection regression. The double residual estimator of Robinson (1988) is obtained by:

$$\begin{aligned} \underbrace{R_i - E[R_i|x_i]}_{u_1} &= (X_i - E[X_i|x_i])'\beta + \lambda(S_i) - E[\lambda(S_i)|x_i] + \varepsilon_i \\ &= \underbrace{(T_i - E[T|x_i])\gamma}_{u_2} + \varepsilon_i \end{aligned} \quad \text{(OA.12)}$$

To avoid imposing any functional form, the estimator replaces the unknown quantities $\mathrm{E}[R_i|x_i]$, $E[X_i|S_i]$ and $E[\lambda(S_i)|x_i]$ by a smooth unknown function estimated by nonparametric (kernel-weighted) estimators. The error term ε_i can be non-normal and assumed exogenous $E[x_i|\varepsilon_i] = 0$. Robinson (1988) shows it is possible to construct root-n consistent and asymptotically-normal estimates from the residuals u_1, u_2 obtained from these nonparametric estimators as: $\hat{\gamma} = (\hat{u}_2'\hat{u}_2)^{-1}\hat{u}_2'\hat{u}_1$. The parameter of interest for x_i can be extracted from $\hat{\gamma}$ without modelling explicitly $m(x)$. To obtain the ordinary inverse Mills ratio, Robinson (1988) shows that if $\lambda(S_i)$ is estimated parametrically (probit), the asymptotic distribution of the estimates is affected unless $\hat{\lambda}(S_i)$ estimated by the nonparametric estimation converges to the estimate from the parametric estimation. To improve the efficiency, we bootstrap at 50 replications of the clustered error terms to account for the possibility that $\hat{\lambda}(S_i)$ does not converge to its parametric estimation. The clustered variance is $V(\hat{\gamma}) = (\hat{u}_2'\hat{u}_2)^{-1}\sum_{j=1}^{n_c}\theta_j\theta_j'\,(\hat{u}_2'\hat{u}_2)^{-1}$, where $\theta_j = \sum_i \hat{u}_i\, t_i$, $\hat{u}_i$ is the residual for the i^{th} observation and t_i is the row vector of T_i; $n_c = 14$ is the number of clusters (districts). We also experiment with different trimming levels incrementally from 0.00 to 0.05. Since the results are similar, we report the default level 0.00.

The general framework of Robinson (1988) allows an extension to account for potential endogeneity of x_i or $E[x_i|\varepsilon_i] \neq 0$. Assume there exists a vector of exogenous instruments Z_i such that Z_i is correlated to x_i but not to ε_i: $x_i = Z_i'\pi + v_i$ and $E(Z_i|\varepsilon) = 0$. Assume that $E(\varepsilon_i|x, v) = \rho\tau$ or $\varepsilon_i = \rho\tau + \eta$. The selection model becomes:

$$R_i = m(x_i) + X_i'\beta + \lambda(S_i) + \rho\tau_i + \eta \tag{OA.13}$$

The partially linear model can be estimated by conditioning on x_i:

$$\begin{aligned}\underbrace{R_i - E[R_i|x_i]}_{u_1} = (X_i - E[X_i|x_i])'\beta + \lambda(S_i) - E[\lambda(S_i)|x_i] + +\rho(v \\ - E(v|x_i) = \underbrace{(T_i - E[T|x_i])\gamma}_{u_2} + \varepsilon_i\end{aligned} \tag{OA.14}$$

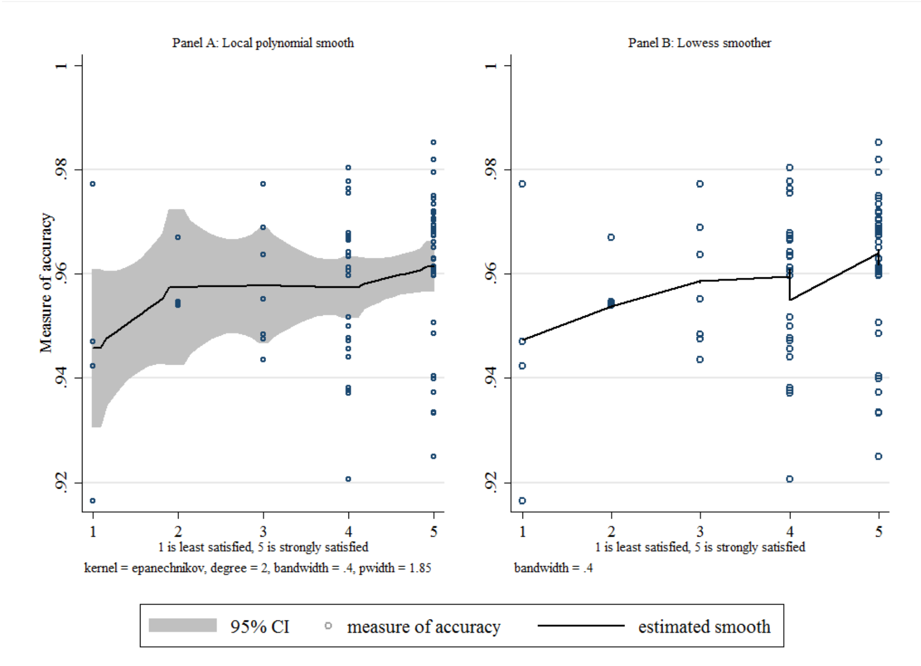
Here, different from the unknown $E[R_i|x_i]$, $E[X_i|S_i]$ and $E[\lambda(S_i)|x_i]$, $E(v|x_i)$ can be parametrically estimated from the residuals of the first stage of IV $\hat{v} = x - Z'\hat{\pi}$. To account for the residuals being estimated, we bootstrap the clustered error terms at 50 replications.

For a semiparametric censored Tobit model of $R_i = \max\{0, X_i'\alpha + \varepsilon\}$, Powell (1984) proposes estimation of the unknown parameters α by the minimiser $\hat{\alpha}_{CLAD}$ (censored least absolute deviations):

$$\hat{\alpha}_{CLAD} = \text{argmin}\ Q_n(\alpha) = \text{argmin}\ \frac{1}{n}\sum_{i=1}^{n}[R_i - \max\{0, X_i'\alpha + \varepsilon\}] \tag{OA.15}$$

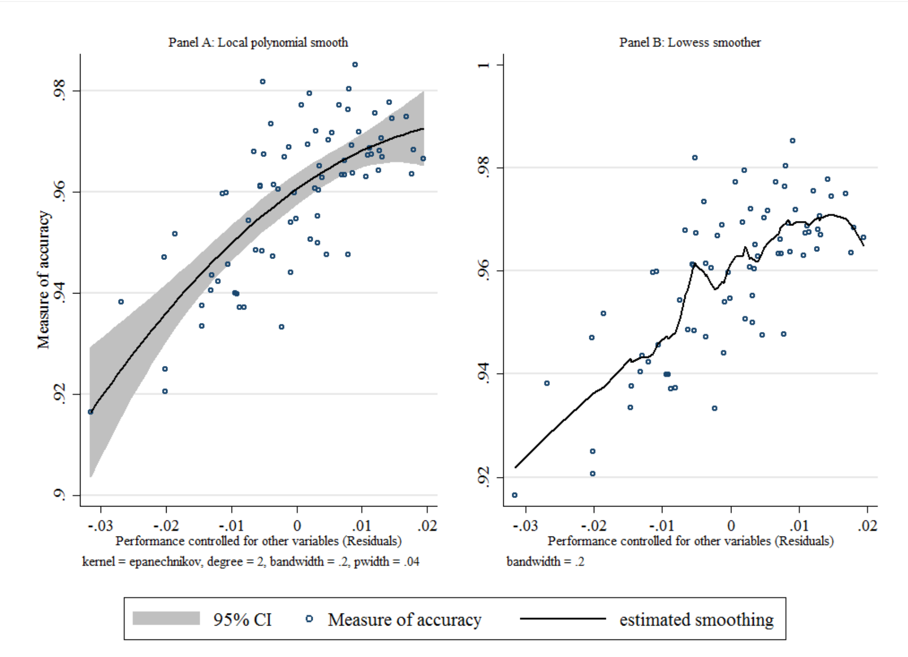
Buchinsky (1994) provides an iterative linear programming algorithm (IPLA) to computationally estimate the parameters.

Figure OA2. Kernel-weighted regressions of report accuracy *(y)* and NGO performance *(x)*



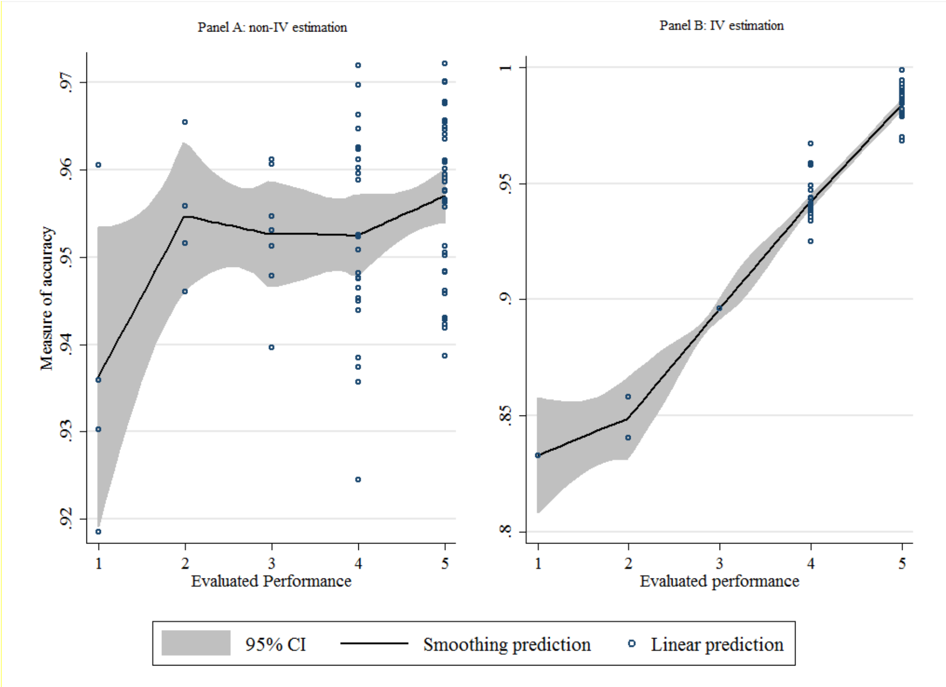
Note. Panel A reports a kernel-weighted local polynomial regression (multiple piecewise linear estimations *locally* in a neighborhood of *x* within a given bandwidth). Polynomials of *x* are included to improve the fit of the estimation. The default Kernel (Epanechikov) distribution and 95% confidence interval bands are used. The direction of the relationship remains similar given reasonable changes of our chosen bandwidth ($N^{-0.5} = 0.4$). Panel B plots the locally weighted scatter plot smoothing, which allows evaluation points near extrema to be downweighted (smoothed using a narrower bandwidth) as in Cleveland and Devlin (1988).

Figure OA3. Kernel-weighted regressions of accuracy measure and performance (residuals)



Note.: We repeat the nonparametric kernel regressions between the accuracy measure (*y*) and residuals (*u*) from an OLS of (*y*) on all control variables except the evaluated performance (*x*).

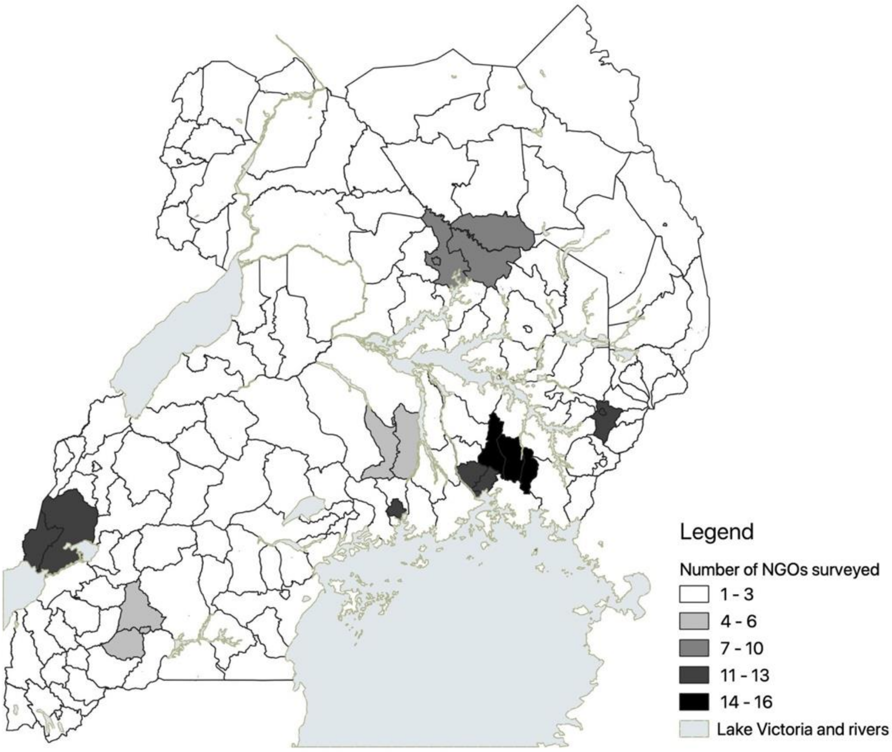
Figure OA4. Semiparametric non-IV and IV estimation following Robison (1988) model



Note. See Online Appendix D for details. IV estimation uses for instruments proposed in Section 6.3. Both Panels exhibit a positive relationship between the measure of reporting accuracy and the evaluated performance.

E. The geography of the surveyed NGOs in 2002

Map OA.1. Districts in the 2002 Ugandan NGO survey.



Note. Darker colors represent more NGOs were drawn from the districts.

F. Additional Tables

Table OA1. IV estimations of selection and outcome equations for control variables

VARIABLES	Selection equation $C_i(.)$	Outcome equation: R_i IV-Heckit	IV-Cragg
Reports requested per year	37.765	-0.455	-0.586
	(104.395)	(0.674)	(0.376)
% Professional degrees	1,604.467**	8.659	0.112
	(723.274)	(7.205)	(3.241)
Registered as company	56.767	12.245***	10.638***
	(388.568)	(3.077)	(2.374)
Received grant (Yes = 1)	982.209**	6.593	1.734
	(423.353)	(5.443)	(4.049)
Lack of funding	-719.620	-0.994	0.679
	(521.026)	(3.709)	(2.423)
Lack of skilled staff	236.746	-0.068	-0.743
	(402.950)	(3.935)	(3.045)
Government as a hindrance	-843.225**	-9.860*	-7.042
	(403.710)	(5.530)	(4.626)
Years working in government	0.313	-0.533**	-0.414*
	(22.731)	(0.222)	(0.224)
Tenure length	-135.144**	-0.697	-0.333*
	(62.618)	(0.546)	(0.195)
Clerical staff	95.982	0.231	-0.024
	(60.820)	(0.259)	(0.117)
Age	208.572*	2.900**	1.502***
	(107.683)	(1.140)	(0.384)
Age^2	-6.471**	-0.088***	-0.043***
	(3.045)	(0.033)	(0.010)
Religious Affiliation	1,712.166***	0.885	-3.013
	(536.460)	(5.019)	(3.213)
Board	-1,114.106	-2.749	0.659
	(704.692)	(6.502)	(4.930)
Member involvement	704.594*		
	(401.664)		
Lambda/sigma		15.209	0.516***
		(11.054)	(0.083)
Constant	-845.115	877.391***	10.042***
	(1,154.951)	(21.426)	(0.679)
Constant	-845.115	877.391***	899.071***
	(1,154.951)	(21.426)	(13.051)
Observations	100	75	100

Note. *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (14) in parentheses (bootstrapped standard errors with 100 replications are largely identical). The set of instrumental variables is not statistically significant in the selection equation. The coefficient θ_2 of the predicted value of the suspected endogenous variable ($\hat{v}$) = -5.962* (3.104), indicating the presence of endogeneity.

G. Robustness to different measures of satisfaction

In the main analysis, the focus group of 6 to 10 participants were asked to rate their satisfaction with the NGO performance by answering the question "the people who live in this parish are satisfied with the performance of [NGO]" on a Likert scale (1 = least satisfied and 5 = most satisfied). In addition, the focus group questionnaire included several other relevant aspects such as: how good the NGO is at what they do (*NGO_good*); how important the NGO is to the community (*NGO_important*); how accessible the NGO is to the community (*NGO_assessible*), and how quickly the NGO responds when the community asks for help (*NGO_quick_respond*). For completeness we report: in Table OA2 the correlations of our preferred measure and the four alternative measures; and in Table OA3 robustness checks using these measures. We note that the alternative measures are positively correlated with our preferred measure and the estimates are generally positive, despite showing less precision. Using other aspects of satisfaction does not alter out main results.

Table OA2. Correlations of alternative measures of satisfaction

	Satisfaction with the NGO performance (1 = least satisfied, 5 = most satisfied)
NGO_good: the NGO representatives are good at what they do	0.58
NGO_important: the NGO is an important part of the community	0.57
NGO_assessibe: the NGO representatives are always available when they say they are going to be	0.05
NGO_quick_respond: The NGO is always quick to respond whenasked for help	0.56

Note. Variables are rated in a Likert scale with 1 = strongly disagree and 5 = strongly agree

Table OA3. Robustness to different measures of Performance

VARIABLES	*MAD statistics*		
	Heckit	Cragg's	OLS
	(1)	(2)	(3)
Performance	3.384***	3.459**	2.916
	(1.282)	(1.648)	(1.846)
NGO_good	3.177*	3.252*	3.067
	(1.779)	(1.744)	(2.125)
NGO_important	1.944	2.008	1.488
	(1.246)	(1.230)	(1.327)
NGO_assessible	1.514*	1.523*	1.610*
	(0.832)	(0.839)	(0.856)
NGO_quick_respond	1.887*	1.995*	1.555
	(1.123)	(1.095)	(1.183)

Note. *Performance* is used in the main analysis, measured as the rating of the beneficiary community on how the people of the parish are satisfied with the general performance of the NGO (1 = Least, 5 = Most). The modified MAD statistics measure the conformity of the reported data with Benford's Law. A larger value indicates a larger deviation from the theoretical Benford distribution. The naïve OLS regression uses the sample with available information on reporting accuracy without correcting for any sample selection; robust standard errors are reported.

References for the Online Appendix

Bettin, Giulia, Riccardo Lucchetti, and Alberto Zazzaro. 2012. "Endogeneity and Sample Selection in a Model for Remittances." *Journal of Development Economics* 99 (2). North-Holland: 370–84. doi:10.1016/j.jdeveco.2012.05.001.

Blundell, Richard, and Richard J Smith. 1994. "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables." *Journal of Econometrics* 64 (1–2). North-Holland: 355–73. doi:10.1016/0304-4076(94)90069-8.

Blundell, Richard W., and James L. Powell. 2004. "Endogeneity in Semiparametric Binary Response Models." *Review of Economic Studies* 71 (3). Wiley/Blackwell (10.1111): 655–79. doi:10.1111/j.1467-937X.2004.00299.x.

Buchinsky, Moshe. 1994. "Changes in the U . S . Wage Structure 1963-1987 : Application of Quantile Regression." *Econometrica* 62 (2). The Econometric Society: 405–58. doi:10.2307/2951618.

Cleveland, William S., and Susan J. Devlin. 1988. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association* 83 (403): 596–610. doi:10.1080/01621459.1988.10478639.

Cragg, John G. 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica* 39 (5). The Econometric Society: 829. doi:10.2307/1909582.

Das, Mitali, Whitney K. Newey, and Francis Vella. 2003. "Nonparametric Estimation of Sample Selection Models." *The Review of Economic Studies* 70 (1). Oxford University Press: 33–58. doi:10.2307/3648610.

Newey, Whitney K. 1987. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." *Journal of Econometrics* 36 (3). North-Holland: 231–50. doi:10.1016/0304-4076(87)90001-7.

Newey, Whitney K., and James L. Powell. 2003. "Instrumental Variable Estimation of Nonparametric Models." *Econometrica* 71 (5). Wiley/Blackwell (10.1111): 1565–78. doi:10.1111/1468-0262.00459.

Powell, James L. 1984. "Least Absolute Deviations Estimation for the Censored Regression Model." *Journal of Econometrics* 25 (3). North-Holland: 303–25. doi:10.1016/0304-4076(84)90004-6.

Rivers, Douglas, and Quang H. Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39 (3). North-Holland:

347–66. doi:10.1016/0304-4076(88)90063-2.

Robinson, P. M. 1988. "Root-N-Consistent Semiparametric Regression." *Econometrica* 56 (4). The Econometric Society: 931. doi:10.2307/1912705.

Sanderson, Eleanor, and Frank Windmeijer. 2016. "A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables." *Journal of Econometrics* 190 (2). North-Holland: 212–21. doi:10.1016/j.jeconom.2015.06.004.

Schwiebert, Jörg. 2015. "Estimation and Interpretation of a Heckman Selection Model with Endogenous Covariates." *Empirical Economics* 49 (2). Springer Berlin Heidelberg: 675–703. doi:10.1007/s00181-014-0881-z.

Semykina, Anastasia, and Jeffrey M. Wooldridge. 2010. "Estimating Panel Data Models in the Presence of Endogeneity and Selection." *Journal of Econometrics* 157 (2). North-Holland: 375–80. doi:10.1016/j.jeconom.2010.03.039.

Smith, Richard J., and Richard W. Blundell. 1986. "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply." *Econometrica* 54 (3). The Econometric Society: 679. doi:10.2307/1911314.

Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57 (2). The Econometric Society: 307. doi:10.2307/1912557.

Wooldridge, Jeffrey M. 2001. "Econometric Analysis of Cross Section and Panel Data." *The MIT Press* 58 (2). MIT Press: 735. doi:10.1515/humr.2003.021.