CREDIT Research Paper

No. 17/10

Better performing NGOs do report more accurately:

Evidence from investigating Ugandan NGO financial accounts

By

Ronelle Burger, Canh Thien Dang, Trudy Owens

Abstract

Improving ways to assess development efforts is an important task. Yet, little has been done to understand the connection between the effectiveness of NGOs and their financial accountability. We use Benford's Law to assess accuracy of financial reports by a sample of Ugandan NGOs. We find 25% of the sample provide financial information that may be inaccurate to the true values. We find NGOs with better beneficiary ratings 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 some requested financial information and the decision to report inaccurately. We find no evidence that the two decisions are related, with the former decision due to limited capacity and skills. The results suggest a bigger role for beneficiary assessments in monitoring the sector, rather than an increasing demand for financial reporting.

JEL Classification: L25, L31

Keywords: NGOs, diversification, risk aversions, motivations, historic flood

Centre for Research in Economic Development and International Trade, University of Nottingham

CREDIT Research Paper

No. 17/10

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

By

Ronelle Burger, Canh Thien Dang, Trudy Owens

1. Introduction
2. The Ugandan NGO survey
3. Using Benford's Law to detect potentially fraudulent NGOs
4. A conceptual framework
5. Econometric methodology and analysis
6. Robustness checks
7. Concluding remarks

Appendices

References

Appendices for Online publication

The Authors

Burger: Stellenbosch University, rburger@sun.ac.za; **Dang**: University of Nottingham, Canh.Dang@nottingham.ac.uk; **Owens**): University of Nottingham, Trudy.Owens@nottingham.ac.uk.

Acknowledgements

We thank Oliver Morrissey for comments on earlier drafts, Julien Labonne, Rafaelle Miniaci, Jan Stuhler for helpful discussions, and participants at OxDev Workshop 2016, German Development Research Group Conference 2017, Nordic Econometrics Meeting 2017 and Nottingham PhD Conference 2017 for comments and feedback. We acknowledge the funding from the World Bank and the Japanese Government for the data collection. Dang thanks financial support from the ESRC [grant number ES/J500100/1], the YITP of the IAAE, and the University of Brescia (Italy) for their hospitality. Usual disclaimer applies. Conflict of interest: none.

1. Introduction

Non-governmental organisations (NGOs) have become an integral part of the international development efforts. NGOs complement local governments by providing public goods and channelling donors' aid to projects in developing countries. In some cases, NGOs have displaced the traditional role of state institutions when donors bypass highly corrupt governments to deliver aid through these non-state channels (Acht et al., 2015). The NGO sector, however, is not immune from accountability (Ebrahim, 2003). There have been requests for an increased level of financial transparency and a crisis of eroded public trust after scandals related to misuses of funding and donations (Herzlinger, 1995; Chen, 2016). Instead of eliciting effort to improve the effectiveness of funded projects, NGO managers have strong incentives to manipulate their financial reports in order to improve their standing and reputation. Previous studies have attempted to construct and validate measures to identify financial misconducts and the underlying mechanisms for corporations (see Dechow et al., 2010) and non-profits in developed countries (Hofmamn and McSwain, 2013). The literature falls short in at least two dimensions. First, there is little research on developing countries, particularly for development NGOs. Second, available measures of reporting accuracy have deficiencies and usually inapplicable to development NGOs whose financial accounts are usually incomplete and lack of forward information.[1] To address these gaps, we use Benford's Law, an easy-to-replicate and cost-effective way, to examine the level of errors in financial data of development NGOs. We study the underlying motivations of their financial transparency by distinguishing between the decision to withhold some data and the decision to report inaccurately when the NGO submits all requested information. Finally, we ask whether financial misreporting matters by investigating the connection of an NGO's performance and the accuracy of their financial data. We revisit the popular belief that the high cost of accounting procedures for monitoring purposes could crowd out serving the client community.

To examine the prevalence of misreporting in financial reports of development NGOs we use a unique representative dataset of Ugandan NGOs (see Barr et al., 2005). Extensive information including financial data, performance, and characteristics was collected through interviews with NGO managers

[1] Prior accounting literature outlines the drawbacks, such as correlation with underlying organisation characteristics and their heavy reliance on forward-looking and time-series data (Amiram et al., 2015).

and focus groups with beneficiary communities. Out of 104 organisations, 26 percent of NGOs failed to provide a complete record of financial information (that is, provided no information, or only revenue or expenditure data despite both being requested).[2] For the remaining 74 percent with complete information, we approximate the degree of inaccuracy in their financial reporting. Following Amiram et al. (2015) and Michalski and Stoltz (2013), we apply Benford's Law to examine the divergence between the distribution of the leading digits of all figures in an organisation's self-reported financial data and the theoretical Benford distribution. Benford's Law posits that a dataset that contains naturally occurring numbers (without conscious manipulation) will have the first digits of all the numbers follow a well-defined, logarithmic, decreasing distribution shown in Hill's (1995) theorem. Greater deviations between the observed data and the theoretical distribution indicate greater degrees of inaccuracy. We employ several methods to measure the convergence of the two distributions, which we use as the quantified degree of inaccuracy in the financial reports. We reject the null hypothesis that the self-reported financial data conform to the Benford distribution for nearly 25 percent of the organisations with complete information. As such, we contribute to the growing literature of forensic economics (Zitzewitz, 2012) and the use of distributional properties of numbers to identify potential misreporting in data that are of the public's interest (see Fang and Gong, 2017 for detection of overbilling in Medicare reimbursements; Michalski and Stoltz, 2013 for detection of falsified economic data at government levels; Almond and Xia, 2017 for detection of manipulation in investment returns of US non-profits).

In addition to inaccuracies in the data provided, there is a second issue that the literature often overlooks, namely that misreporting may result from the deficiency of an organisation's reporting process rather than their management's intent. Unlike corporations whose accounting and finance departments are sizable, development NGOs face trade-offs between charitable spending and administrative expenses. As such, financial misreporting could arise without the NGO's intention. Woodwell and Bartczak (2008) report that 80% of surveyed donors did not provide sufficient overhead allocations to cover the time and expenses their recipients incurred on reporting requirements. We ask whether non-disclosure is primarily due to unavailable financial accounts (plausibly related to a

[2] Non-disclosure incidence is well documented in both the corporate literature (see Leuz and Wysocki, 2016) and in the non-profit sector in developed countries (Johnson and Prakash, 2007; Hofmann and McSwain, 2013).

shortage of resources or financial skills) or deliberate concealment of available financial accounts. This distinction is statistically important because it affects how we treat the missing observations. There are several applications of this design in labour economics (see Bettin et al., 2012 on remittances) and health economics (Dow and Norton, 2003; Madden, 2008). We are not aware, however, of any such application in the information asymmetry literature in distinguishing between two types of information senders. The first type are those whose decision to provide full requested information is uncorrelated with the latent decision to report information accurately: these senders would always provide full information if available, with incomplete information arising from *ex-ante* factors unrelated to the latent accuracy. The second type are those whose decision to provide incomplete information arises from factors related to the latent information accuracy: they only provide full requested information if the optimal accuracy of their reports exceeds a reservation level. We answer this question by comparing the model fit of a Tobit specification (exogenous sample selection when missing observations are genuinely not available) and a Heckman specification (endogenous sample selection when missing observations are self-selectively withheld). Using the Ugandan NGO dataset, we estimate a Cragg's (1971) double-hurdle model and a Heckman sample selection model and test for the preferred model using the Vuong (1989) non-nested test and a test proposed by Silva et al. (2015).

In investigating whether non-disclosure of all information is motivated by similar factors as provision of inaccurate information, we find the double-hurdle model outperforms the Heckman-type selection model, suggesting that the decisions appear to be uncorrelated. Further analysis suggests that the lack of human resources could be a significant constraint towards transparency in the Ugandan NGO sector. NGOs with more clerks and a larger proportion of staff having a degree have a significantly higher propensity to provide full financial information once requested. NGOs lacking human resources or skills may choose to keep only information necessary for their operations, resulting in the provision of incomplete financial data. The evidence suggests that in Uganda the prevalence of financial reporting inaccuracy could perhaps be driven by the lack of resources dedicated to accounting tasks rather than manipulation to hide mediocre performance.

Finally, we contribute to the literature on the impact of performance-based compensations on misreporting (see Burns and Kedia, 2006) by exploring the relationship between performance of the

NGO as evaluated by their direct beneficiary community, and accurate financial reporting to a third party. NGOs face a trade-off between the high cost of accounting procedures for monitoring purposes (referred to as upward accountability) and serving their community (downward accountability). If reporting accuracy, measured as deviations from Benford's Law, is aligned to NGO performance, it could help as a signal to detect NGOs that are underperforming or not serving the community well. Our results indicate that NGOs that are poorly evaluated by their communities tend to provide less accurate financial information. The results are robust when we relax the distributional and functional assumptions underlying Cragg and Heckman specifications by using non-parametric and semi-parametric alternatives. To ascertain that the mechanism is not purely caused by some unobserved motivations and altruism that lead to both higher beneficiary satisfaction and more accurate reporting, we instrument for the evaluated performance variable using two-stage Limited Information ML (LIML) estimators for the Cragg and Heckman models with endogenous variables. We exploit the independence of the composition of the focus groups who evaluate the NGO's performance to construct the set of instruments. We use the demographic characteristics of the community respondents, namely the percentage of senior-age participants and the percentage of participants with a connection with the evaluated NGO. These characteristics are strongly correlated with their evaluation scores, but are irrelevant to the NGO's decision to provide accurate financial information to a third party (unknown to the community). We find the positive relationship between the degree of inaccuracy in the financial information and the NGO's perceived performance is robust. The result shows that organisations with stronger performance may have fewer incentives to hide information and be more willing to provide accurate financial data to a third party or an external body.

The paper is organised as follows. We describe the data in Section 2. Section 3 describes Benford's Law and its application to our survey 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. Section 7 concludes.

2. The Ugandan NGO survey

We match data from two sources in a 2002 survey of the Ugandan NGO sector.[3] The first survey module (NGO questionnaire) was administered to a random sample of NGOs in 14 districts of Uganda (see Online Appendix E) drawn from a verified NGO Registration List held with the Ministry of Planning within the Ugandan Government (details of sampling are available in Barr et al. 2003). The face-to-face interview with an NGO representative (usually the head of the NGO) was conducted by Ugandan field workers.[4] The questionnaire asks 255 questions, including information regarding the organisation's expenditure and income (in 2001 and 2002), funding, governance and activities. Questions regarding revenues and expenses were designed such that the information was highly aggregated and should have been readily available from a standard annual account.[5]

The second survey module captured community needs, characteristics and activities via a structured focus group interview. 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.[6] The enumerators contacted the parish leaders asking them to recruit between six to ten community members for the group meeting. The selection process ensured comparability and consistency of the community evaluation across NGOs. The interview assessed how beneficiary communities evaluated the NGOs surveyed in the first module. Focus group participants were asked to rate their satisfaction with the NGO on a Likert scale (1 = least satisfied and 5 = most satisfied). Various characteristics of the focus group participants were also collected.

[3] We compliment the 2002 survey with financial information from a follow-up survey in 2008 conducted by Burger and Owens (2010). As the 2008 survey provides less information, we base our analysis on the 2002 wave.
[4] 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. The wording reads as follow: "The Government of Uganda appreciates the work that NGOs do and the services they provide to the people of Uganda. Unfortunately, there is a lack of knowledge and a lot of confusion regarding the NGO sector. The purpose of this survey is to dispel confusion and improve knowledge regarding the NGO sector in Uganda. To be successful, we need your help. The ultimate objective of this study is to enable the Uganda government and international donors to assist better NGOs operating in the country. Please be assured that the information collected will be used for research purposes only and will be treated as confidential."
[5] If the respondent was unable to answer the questions, either the NGO representative was asked to request the information from a colleague or the enumerator left the relevant section of the questionnaire to be completed and collected on a designated day that suited the NGO. In larger NGOs, the enumerator gave the representative the relevant section before the interview so that the NGO could prepare the required figures.
[6] The enumerators in the focus group interview used a series of questions to filter out focus groups that did not know about the NGO working in their community.

While we aimed to ensure that the characteristics of the focus group participants were orthogonal to the surveyed NGO by never involving the NGO in the selection of the surveyed community or any other aspect of the evaluation, one caveat is necessary. Some of the listed NGOs might have favourable connections with the leaders of the parishes, who could then select focus group participants that could give biased evaluations of the NGOs. In practice, we believe this was not the case. Since the parish leaders were not told the name of the NGO that the group would evaluate before selecting the meeting participants, it is unlikely that the leaders would be *ex-ante* able to strategically recruit the focus group that might have a strong bias favouring the NGO. Anecdotes from the interviews support the argument. In some cases, the enumerators needed to prompt the group about the NGO being evaluated.[7] Barr and Fafchamps (2006) provide further description of the second survey module.

We match the NGO surveyed in the first module with the respective community from the second module.[8] During the data compiling process, we discovered four enumerators had exhibited "cheating incidents" while conducting the NGO interviews.[9] 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 financial accuracy of NGOs

The accounting literature focuses on measuring earning managements in accounting reports of corporations (see Dechow et al., 2010). Methods include accrual-based estimates from the Jones models (Jones, 1991), or distributional analyses (Burgstahler & Dichev, 1997). Rather than focusing on accruals and earnings, non-profit studies often attempt to measure irregularities in reports by estimating either expected programme ratios (Trussel, 2003), levels of charity care (Vansant, 2011) or fundraising and administrative expenses (Yetman and Yetman, 2011). There are, however, several weaknesses inherent in both approaches. Within the accounting literature, first, measures estimated from prediction models

[7] There is one group whose participants were all connected to the evaluated NGO. Excluding this group does not alter the finding.
[8] 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.
[9] It is a common problem in collecting survey data (see Judge and Schechter, 2009).

incur sample selection bias and measurement errors (Dechow et al., 2003). Second, these measures require strong assumptions about the organisation's objective function and managers' incentives, which are not always realistic and could induce correlation between the measures and the organisation's characteristics (see Amiram et al., 2015). Third, these models require forward-looking information 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 (Hofmann and McSwain, 2013). With respect to the non-profit studies the focus on potential errors in some specific categories ignores the fact organisations could manipulate the data as the whole.

We use an alternative proxy for measuring the accuracy of a self-reported set of financial data based on Benford's Law. It is a mathematical law regarding the frequency distribution of leading digits in many sets of numerical data (e.g., the leading digit of the number 1201.17 is 1). Contrary to the basic intuition, the occurrence of each digit as a leading digit is not equally likely (uniform). Instead, Hill's (1995) theorem states that if distributions are non-truncated or uncensored, random samples of varying magnitudes taken from a random mixture of those distributions will have the first digit converging to the logarithmic of a distribution, dubbed the Benford distribution (See Appendix A).

In economics, Hal Varian (1972) first suggests Benford's Law as an effective tool to uncover irregularities in accounting and financial data.[10] The premise is that the distribution of the first digits of all figures in an accurately reported financial statement is expected to conform to the Benford distribution. Those who fill in the reports and make up figures may be biased towards simpler or more intuitive distributions such as the uniform distribution. The biased report is then expected to deviate from the Benford distribution. The biases are captured by measuring these deviations from the theoretical distribution. Durtschi et al. (2004) outlines four requirements for applying the law to detect frauds in 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 positively skewed with a median that is lower than the mean. Lastly, there should be no numbers that result from a single

[10] Some notable applications include Judge and Schechter (2009) to detect fieldworker cheating in household surveys; Muller (2011) to detect financial misreporting of Greek financial data to ensure entry into the European Union; and the detection of falsified voting in the Iranian 2009 elections (Battersby, 2009).

multiplication of numbers (e.g., quantity × price). Our NGO data satisfies all these criteria. As such, Benford's Law is particularly appealing and the only viable method for the type of data often available for developing country NGOs. It is also an easy-to-replicate way of evaluating the level of reporting errors within underlying data, providing a cost-effective way to flag potential misreporting behaviour on a large scale.

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, 2011). Following Amiram et al. (2015), we use the Mean Absolute Deviation (MAD) statistic in the main analysis, and report other popular measures derived from Benford's Law in the Appendix C. 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 proportion according to Benford's Law (see Table 1):

$$\text{MAD} \equiv \frac{1}{9}\sum_{i=1}^{9} |P_o(d_i) - P_e(d_i)| \tag{5}$$

where $d_i = 1, 2, \ldots, 9$ represents the first digit; $P_o(d_i)$ is the observed frequency of digit d_i, and $P_e(d_i)$ is the expected frequency of digit d_i according to Table A1 in the Appendix. Nigrini and Mittermaier (1997) show that since the MAD statistic is independent of the pool of digits used, it becomes preferable to the other proxies when examining larger pools of digits and comparing deviations of financial statements across organisations with different numbers of non-zero financial items reported. Also, as there is no critical value involved in comparing the MAD across organisations, the statistic provides a clear and objective measure: 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 and manipulation. 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 sets of financial statements.

We construct the degree of reporting accuracy of NGOs who do not provide all requested figures as zero to allow for estimating the Cragg model (discussed in the next section). Our dependent variable of interest is: [11]

$$R_i = \begin{cases} 1 - \text{MAD} & \text{if } C_i(.) = 1 \\ 0 & \text{if } C_i(.) = 0 \end{cases} \quad (6)$$

where higher R_i indicates that the complete report is more accurate. The number 1 is chosen arbitrarily to construct a measure of accuracy as the MAD measures the level of reporting inaccuracy.

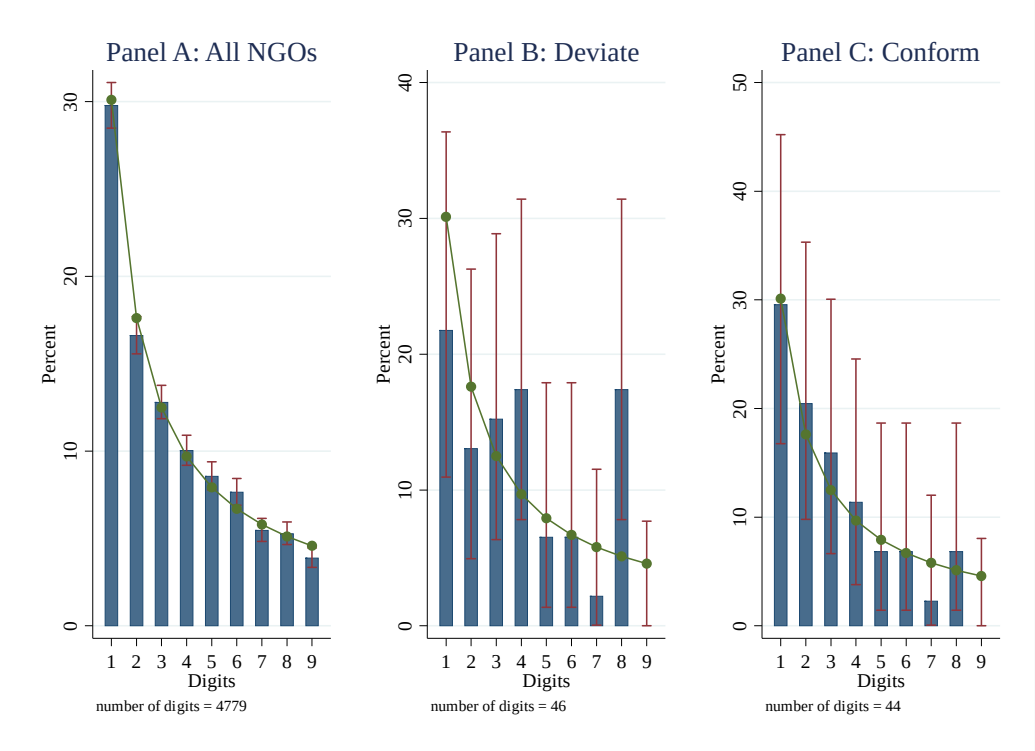
Since there is no critical value associated with the MAD method, it could become impractical to flag potential misreporting. We employ the conventional Pearson's Chi-square test for fit of distribution to test the hypothesis that the observed distribution of all the first digits follows the Benford distribution.[12] We use several alternatives in Section 6. Figure 2 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, further assuring our premise that accounting data should follow the law.[13] When an individual NGO is considered, we reject the hypothesis that the observed distribution of the first digits of all the numbers follow the theoretical distribution for 25% of the sample with complete disclosure. We call these NGOs "deviate" from the law, suggesting that these organisations' full financial accounts may contain inaccuracy detectable by the law. Panel B provides a representative distribution of this group. In contrast, we fail to reject the null hypothesis for the remaining 75% of the sample with complete disclosure. We call these NGOs "conform" to the law, indicating that we fail to find evidence of potential misreporting. Panel C provides an example. We explain the conformity in the next sections.

[11] 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.

[12] Morrow (2014, working paper) and Barabesi et al. (2016) provide discussions for the usefulness of the conventional tests for fit of distribution.

[13] 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 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.

Figure 2. The Ugandan NGO financial data and conformity to Benford's Law



Notes: Bars represent the theoretical frequency of digits 1 to 9 appearing as the first digits according to Benford's Law. Lines represent the observed distributions in three samples. Capped spikes represent confidence intervals at the 95% significance level. In Panel A, we lump all the numbers in all financial accounts provided by the NGO. 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 Pearson's Chi-square test. Panel C is a representative NGO (75% of the sample) which we fail to reject the hypothesis for its requested financial accounts. *Source:* Authors' analysis based on the 2002 Ugandan NGO survey data.

4. A conceptual framework

We review previous theoretical papers to conceptualise factors behind reporting behaviour of an agent: whether and how accurately they reveal their financial information. A formal model is in the Online Appendix OA.

4.1 Why do NGOs provide incomplete financial information?

Despite the surveys being designed to make providing financial information easy, several NGOs, unsurprisingly, 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 financial items for both 2001 and 2002 fiscal years. There are 77 such NGOs in the sample. We define NGOs as having *incomplete disclosure* if they fail

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. Table A2 provides several descriptive statistics on these subgroups.

We aim to understand the mechanisms driving these non-disclosure incidents. This is vital for our empirical analysis which investigates the accuracy of the NGOs' self-reported information. 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 possess all the requested financial information but choose (strategically) to provide the enumerator with none or only a part of their data; other NGOs may simply not have 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). Verrecchia (2001) reviews the literature and provides a theoretical taxonomy for each case, namely the discretionary-based disclosure and the efficiency-based disclosure.

The discretionary-based disclosure explanation postulates that a sender (in our case the surveyed NGO) observes private information about the true state of their financial situation and strategically communicates to a receiver (donors, the public, or the enumerator) at their own discretion.[14] As such, there is a situation where incomplete disclosure can be optimal even when the requested information is available, particularly when there are several *ex post* costs associated with the information to be disclosed.[15] First, the intermediate cost of an unfavourable report may outweigh the credibility gain (see Teoh and Hwang, 1991). The information disclosed could reveal the human capital of the NGO leader, competence or managerial incentives, or the organisation's inefficiency (see Kothari et al., 2009). Second, the NGO could incur costs to clear "doubts in the minds of the uninformed" (Verrecchia, 2001).

[14] Hofmann and McSwain (2013) note that non-profit organisations are not legally required to provide financial statements except to certain government agencies, such as Form 990 to The Office of Management and Budget in the US and the Charity Commission in the UK. NGOs in many developing countries, for example Uganda, face no such enforcement mechanisms but are required to submit tax returns for tax exempt purposes.

[15] Li et al. (2016) study another cost when the information revealed to aware customers spills over to some initially unaware customers. The information then further reduces the perception of the initially unaware customers about the organisation's performance. Note that we exclude costs associated with information acquisition, which we discuss in the efficiency-based approach.

For example, an NGO that revealed unusual expenses might need to exert resources to justify the spending to their beneficiaries or donors. Third, as suggested by experimental studies, (see Gneezy et al., 2013), an altruistic NGO who is incentivised to disclose incorrect information may incur costs due to an intrinsic aversion to lying. Communication theory suggests that in such cases the NGO may strategically withhold some financial figures, despite having access to a complete set of the requested information. The intuition is that the NGO knows if it provides all requested information the optimal level of the information's accuracy would have to be too low, lower than a reservation level (denoted $\underline{R}$ below), in which case the optimal option is to withhold the information in the first place.[16] Such strategic incomplete disclosure causes a problem of data observability: some NGOs self-select to only reveal partial information with the rationale related to the latent level of reporting accuracy.

The efficiency-based disclosure explanation suggests that non-disclosure is an efficient choice concerning resource allocation. Without prior knowledge of the information (good or bad) the organisation chooses to keep an incomplete set of information regardless. The logic is that if the organisation 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 organisation *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 financial information that is necessary and productive to their operation. As such, the decision whether to provide *complete disclosure* is taken before (and independently 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 provides all *available* information during the interview regardless of the possibility of being incomplete.[17] In this case, the incomplete disclosure observed is not a strategic communication to withhold information, but rather a corner solution to a maximisation, subject to some *ex-ante* constraint.

[16] The reservation level could be just the intrinsic aversion to lying. If the NGO knows that complete disclosure forces them to choose some optimal degree of misreporting that is too high (e.g. not worth lying), the NGO decides to withhold some information.

[17] Another explanation is that the NGO has no incentive to reveal their financial situation with 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

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 prediction is that high-performance organisations have incentives to disclose more information to differentiate themselves favourably from other organisations with "bad news" and thus avoid the problem of adverse selection (Spencer, 1973; Dye, 1985; Verrecchia, 1983). 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 information more accurate to their current state. Meanwhile, NGOs who are underperforming would have incentives to manage their figures to increase the possibility of earning new grants, which often are tied to their performance (a similar argument to Healey's (1985) analysis of bonus scheme and corporate management). Another literature suggests that "good" organisations also have incentives to issue detailed and accurate records of their financial situation to avoid potential punishments associated with being detected as untruthful, which could "create doubt about the true motive for which good deeds are performed" (Benabou and Tirole, 2006). We formalise this prediction in the alternative form (the implicit null hypothesis is that there is no significant relationship).[18]

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 formalise this alternative prediction as:

heterogeneity of the enumerator, such as NGOs responded differently to different enumerators, we include dummies of the enumerators' IDs and receive similar results.

[18] Lang and Lundholm (1993) show firms with superior upcoming earnings performance have a higher disclosure propensity to reveal their good news and have a higher score for disclosure quality. Miller (2002) documents a strong positive relationship between earnings announcements and the quantity, revenues and types of disclosure by firms. Lee et al. (2006) provide evidence that earning quality, defined as the proportion of true economic earnings in total reported earnings, increases with earnings performance. Clarkson et al. (2008) a positive association between environmental performance and the quality of environment disclosure.

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. In both situations, the incomplete disclosure is the outcome of a completely free choice, although the mechanism underlying each is different. While the decision to withhold under (i) may be correlated with the consequent degree of report accuracy, it is safe to assume that the data unavailability under (ii) is independent of the decision governing the subsequent report accuracy. Once the NGO provides all requested information, we hypothesise that NGOs with better performance would have an incentive to report more accurate 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 organisations fall short of their financial report quality.

5. Econometric methodology and analysis

5.1 Distinguishing between deliberate deviations from Benford's Law and unintended inaccuracies

Since we are only able to construct measures of reporting accuracy for NGOs with full information, a non-negligible portion of the surveyed sample does not enter our primary estimation, causing a possible sample selection bias. We therefore examine whether withholding incomplete information occurs exogenously or endogenously. As reviewed in Section 4, an NGO can engage in two actions. First, it decides on whether to provide the enumerator with all information requested (both expenditure and revenue related figures). Second, if it does provide a full account, the NGO decides on the level of accuracy of the account. Formally, let $\mathrm{C}_i(.)$ be a binary function for NGO i such that $\mathrm{C_i}(.) = 1$ if i provides all requested information in the first stage, 0 otherwise. Let $\mathrm{R_i}$ be the measure of reporting accuracy, which was specified in Equation 6, of NGO i once we observe all requested information in the second stage. 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:

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

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

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 assumptions in the first stage: 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 information and how accurately to report are governed by two independent mechanisms. The intuition is that incomplete information here is a corner solution caused by either a budget constraint (gathering full information is *ex-ante* financially infeasible) or 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 $cov(v_i, u_i) = 0$. As such, incomplete disclosure is a genuine observation. We assign the value of accuracy for the NGO with incomplete information as having zero-accuracy. The appropriate model would be a Tobit-type with censoring point at $R^* = 0$. Practically, we use the Cragg (1971) double-hurdle model for censored responses to allow for two independent mechanisms underlying the selection and the outcome decision. Another attractive feature is that homoscedasticity and normality conditions are not necessary for consistency of the estimator (see Wooldridge, 2010).

$$R_i^* = \begin{cases} R_i & \text{if } C_i(.) > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{(Tobit model)} \quad (3)$$

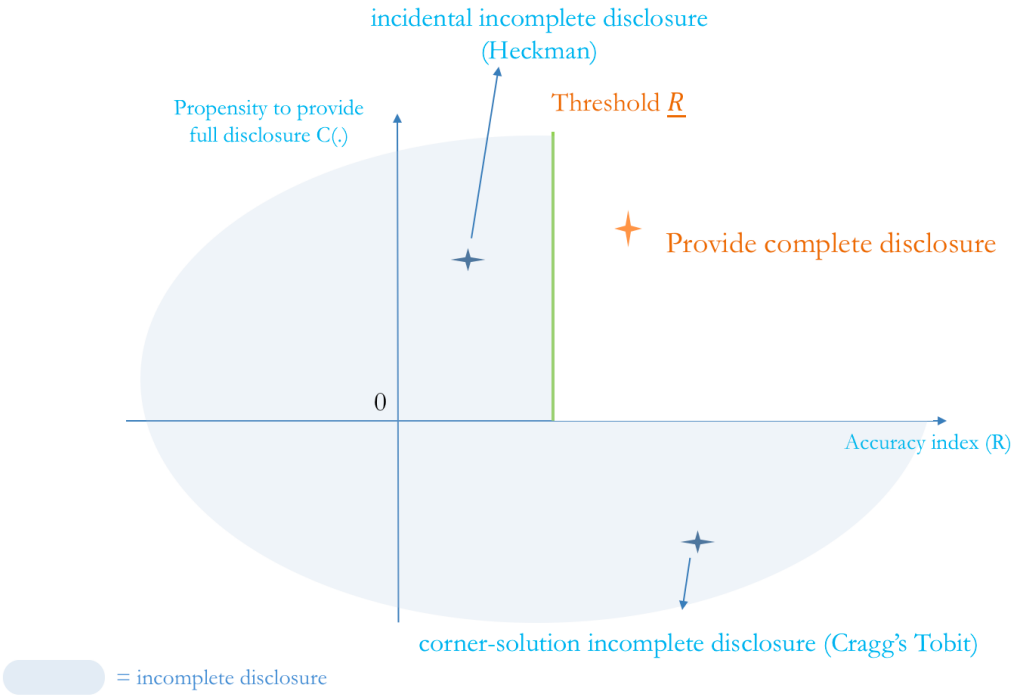
In contrast, if the incomplete information is an *ex-post* strategic decision, the two stages are governed by two related mechanisms: NGOs after considering their level of reporting accuracy at the second stage would decide on whether to reveal full information (endogenous sample selection on the dependent 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 a 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: $cov(u_i, v_i) \neq 0$. As such, 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 together. Different from the former, the reporting accuracy of the latter are available but unobservable to the researchers because of the intention of the NGO. We assign the value of reporting accuracy for these organisations as latent observations, instead of genuinely zero. The appropriate model would be a Heckman selection model. A similar method was used in Burger and Owens (2010) exploring the transparency of these NGOs.

$$R_i^* = \begin{cases} R_i & \text{if } C_i(.) > 0 \text{ and } R_i^* \geq \underline{R} \\ NA & \text{otherwise} \end{cases} \qquad \text{(Heckman model)} \qquad (4)$$

In sum, we have three categories of NGOs: (i) an incomplete disclosure (corner-solution) group who do not keep a complete financial record due to some institutional constraints (such as fixed costs of information gathering); (ii) another incomplete disclosure group who keep a complete financial record but self-select to withhold some information; and (iii) a group of organisations who keep a complete financial record which they provide to the enumerators. Figure 1 illustrates the three categories.

Figure 1. Three types of NGOs regarding information disclosure and report accuracy



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

Like Dow and Nortons (2003) and Bettin et al. (2012), we aim to identify the appropriate model that fits the data statistically, through which we then infer the dominating selection mechanism explaining the incidence of incomplete information. The standard approach is to check the significance of the inverse Mills ratio coefficient generated from the Heckman sample selection model. However, Norton et al. (2008) and Silva et al. (2015) argue that such practice does not give reliable information about the ability of the model to describe truncated response data. As such, we abstain from the use of the inverse Mills ratio as our main tool.[19] Instead we use Vuong's (1989) non-nested hypothesis LR test (see Online Appendix B) and a regression-based specification test developed by Santos Silva et al. (2015), based on Davidson and MacKinnon (1981).[20] If the Cragg's model is preferred, we conclude that 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 model

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

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

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

where R_i is the accuracy measure from Benford's Law, $Performance_i$ is the evaluation of NGO *i*'s services by the respective community from the second survey module ranked on the Likert scale (1 = least satisfied and 5 = most satisfied).[21]

Z_{2i} is the exclusion restriction to ensure the consistency of Heckit estimates. We use a binary variable of whether members of the NGO need to vote before the organisation introduces any new activity or service (*Member involvement* = 1 if Yes, = 0 otherwise). The variable proxies for the involvement of members in the organisation's monitoring activities and is expected to satisfy the

[19] Dow and Norton (2003) instead propose an adjusted empirical mean square error test, which is computationally more difficult to implement.
[20] We implement the test using command –hpc-- in Stata, provided by Silva et al. (2015).
[21] We also estimate a reduced selection equation when $\text{Performance}_\text{i}$ is explicitly excluded so that other variables, including instruments, absorb its effects. The results are similar.

exclusion restriction. Stronger member involvement, reflecting the power to veto any new activity introduction, may create incentives or pressures for the organisation to be transparent about their financial activities. These organisations might be more likely to prepare complete records of their financial situation in order to present their case to members for the introduction of a new service. Yet, there is no reason why the organisation would provide an accurate recording response to their members' involvement in their governance. In fact, Olken (2007) uses randomised control trials to show that grassroots participations of the beneficiary communities in monitoring projects had little impact on corrupt behaviour of the contracted agents. We expect that *Member involvement* is a significant explanatory variable of the selection mechanism, but can be excluded from the main outcome equation.

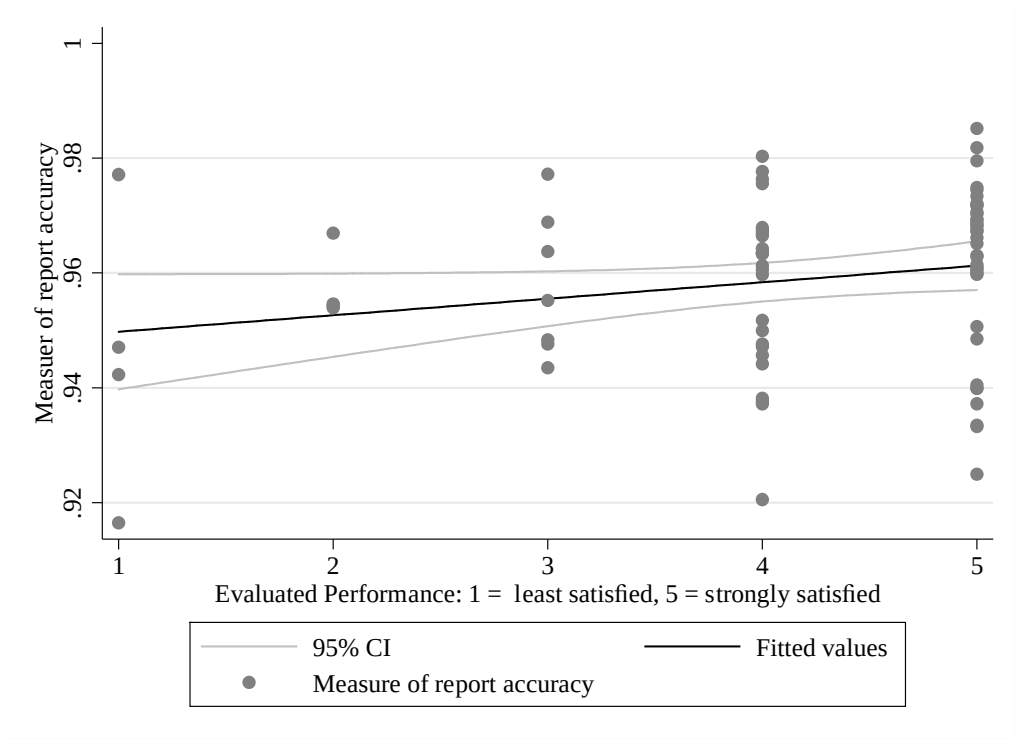
X_i is a vector of NGO characteristics identified in the existing empirical literature as possible explanatory variables of information disclosure and reporting accuracy. We divide the control variables into two groups: variables that proxy for the benefits and cost of transparency common to any organisation, and variables specific to information disclosure in a development context. Appendix B provides a full account of these variables.

Since the survey was conducted in 14 districts, intra-group observations might share common characteristics that may affect the decisions regarding disclosure and accuracy (for example, a local social norm). Including dummies for 14 districts is not a viable option because it 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) since the results are qualitatively similar without clustering.

5.3 Descriptive statistics

The number of NGOs who provided information on revenue and expenses was 77 (74% of the sample). Of these, 58 conform to the Benford distribution of first digits using the Pearson's Chi-squared test at the 5% confidence level. The average performance of these NGOs was higher than the 19 NGOs whose financial figures did not conform. There is no significant difference in performance between the 77 that provided all requested figures and the 27 that did not (see Table A2 in the Appendix). By way of preview, Figure 3 suggests a positive relationship between the evaluated performance and the level of accuracy in the NGO self-reported financial figures.

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



Notes: 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 focus groups selected by the community leader.

5.4 Complete disclosure and determinants

Table 1 presents determinants of an 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 results for the disclosure mechanism from a logit first-stage estimation are as expected. Having an oversight board of directors or trustees is negatively associated with the propensity to provide all requested information to a third party. One potential explanation is the "unintended chilling effect" suggested in Cormier et al. (2005). The oversight board may place lesser importance on a full record of standard financial figures, and the NGO subject to the board agreement may have decided to keep only necessary data.[22] Being registered as a company is not significantly associated with higher propensity to complete disclosure, raising a similar question of effective monitoring by responsible ministers in the Ugandan government office in early 2000s (Deininger and Mpuga, 2005).

[22] Thakor (2015) provides the same argument in a study on the relation of voluntary disclosure and agreement between a firm and investors on the importance of disclosure.

Having received a grant in the past is a significant positive predictor of complete disclosure. Previous studies posit that organisations that need to facilitate fundraising activities have a higher propensity for information disclosure (Jensen and Meckling, 1979). As such, a transparent financial record may serve as a signalling vehicle to attract funding. Having a religious affiliation is a positive predictor of complete disclosure, replicating previous studies.[23] Regular reminders of moral codes through religious preaching is seen to be effective in promoting transparency.

We also find significant effects of reputation loss and manager's career concerns on the propensity to complete disclosure. There exists a significant U-shape effect of the organisation establishment proxied by Age and Age^2. More established NGOs tend to conform to the sectoral norm in financial reporting. When they become sufficiently established, the organisations become less attentive in keeping standard, complete information. One explanation could be the "unintended chilling effect" suggested above. The most established NGOs may become complacent over time and be more relaxed with the sectoral standard. Regarding career concerns, there also 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 usually values transparency. These managers are also the most familiar with the NGO operations and may decide that selectively keeping necessary information may be an efficient strategy for their NGOs.

We find further supporting evidence that human resources may be a significant constraint towards transparency in the Ugandan NGO sector. We find no statistical evidence that worse performing NGOs withhold requested information, suggesting that it is not the performance that incentivises the NGOs to hide information from a third party. Instead, NGOs with fewer clerks and a smaller proportion of staff holding degrees are significantly less likely to provide all requested information. Organisations endowed with a smaller qualified workforce may be constrained and possibly reluctant to exert resources to keep all the standard financial data. NGOs whose managers consider government as a

[23] Mazar et al. (2008) show that individuals are less likely to dishonestly report performance for personal gains when occasionally reminded of moral codes of conduct. Similarly, Dyreng et al. (2012) and McGuire et al. (2011) find that religion-affiliated organisations to be less likely to engage in financial reporting irregularities and more open to information disclosure.

hindrance to their daily operation are less likely to provide all requested financial information. Managers may respond to the lack of human resources and transparency burden by selectively recording only figures they see as necessary to the operation of the NGO. The managers also seem to respond to their member activeness regarding financial transparency. We find that enabling members to vote for the introduction of new services is strongly correlated with the higher propensity to provide all requested financial figures. We conjecture that the deep involvement of the membership system might put pressure on management to be more transparent and accountable. The significance of *Member Involvement* also ensures the exclusion restriction for the Heckman model.

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. NGOs who claim lack of skilled staff and funding as major constraints to their operation also do not have significantly different propensity to provide all requested information. There are three explanations for the insignificant results. First, although the signs of the variables are as expected, the small sample size may well lead to the insignificance (the confidence intervals for these estimates may have become so large that they include zero). Second, there may be biases due to endogeneity. One piece of evidence to support this explanation is that under specifications with endogenous regressors in Table A4, claiming lack of skilled staff as a major constraint becomes significantly associated with a lower propensity to be transparent. After controlling for endogeneity, the result aligns with the above discussion on human resource constraint. Third, we may interpret these insignificant results as potential evidence to support our hypothesised censoring mechanisms. If a Heckman censoring occurs, the NGO representatives may have incentives to withhold figures during the interview even when they have complete records of their financial situation following their donor requests. This strategic behaviour could offset the potential positive effect of donor requests on the propensity to be transparent. If a double-hurdle censoring occurs, other proxies for resource constraints may already have absorbed any significant effect of heavier information gathering costs due to more frequent donor requests. Distinguishing the censoring mechanisms is necessary to explain the insignificance.

Table 2 provides Vuong's (1989) and the HPC (Santos Silva et al., 2015) model selection tests for the nature of the disclosure mechanism. The statistics unanimously indicate that the double-hurdle

model fits our data better than the Heckman sample selection model. Statistically, the mechanism underlying some *incomplete disclosure* is corner-solution censoring: controlling for potential observed determinants, any unobservable confounders underlying the decision to disclose and the decision of report accuracy can be independent. We reject the hypothesis that there exists some reservation level of report quality such that the NGO would withhold the financial figures if their data accuracy fails to exceed the reservation level. Section 4 suggests it is the fixed cost of information gathering that plays a role in the disclosure strategy, supporting the discussion on human resource constraints above.

Although we statistically support that a corner solution censoring mechanism fits the Ugandan data better, the two models generally provide similar descriptions of the significant determinants in the main outcome equation. Statistically, the estimated magnitudes are similar with only slight differences in standard errors, regardless of whether the presence of some fixed costs of information gathering (Cragg's model) or the reservation level of report accuracy (Heckman model) is considered as *a priori.* As a result, we mainly refer to the estimates of the main equation from the double-hurdle model in the next section, noting that the discussion is also valid for the Heckman model.

Table 1. 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)
Oversight 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)
Age	139.149***	1.093	1.181***
	(51.613)	(0.681)	(0.381)
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	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

Notes: *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (14) in parentheses. 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. Appendix B provides a description of the controls.

Table 2. Model selection tests

The HPC test for the preferred model (Santos 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

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

5.5 Report accuracy, NGO performance and other determinants

We support the hypothesis that NGOs with higher evaluated performance provide financial data that are more accurate, 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 this result as further support for a double-hurdle censoring model: two independent mechanisms govern the two decisions. While the disclosure policy may be subject to costs and human constraints; conditional on complete disclosure, NGOs with better charitable performance provide data that are more accurate, perhaps to reveal their performance type (Verrecchia, 1983) or to avoid punishments associated with being detected as untruthful (Benabou and Tirole, 2006).

The coefficient of other NGO characteristics on the degree of report accuracy aligns with our predictions. Having registered as a company is positively associated with the financial data being more accurate, possibly due to regulatory effects or interactions with other companies that are more experienced in accounting. A U-shape relationship also emerges between the organisation establishment and the accuracy of their data. More established NGOs are associated with higher degrees of report accuracy, while the most established organisations are associated with lower accuracy. Besides the complacency explanation, the scale of bookkeeping in larger organisations 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). In fact, we do not find any significant effects of having more clerks and staff with degrees on the organisation's report accuracy. Although having a larger workforce increases the propensity to complete disclosure, more human resources are not statistically associated with higher

report accuracy. Consistent with anecdotal observations in the Ugandan NGO sector, we conjecture that lack of training may be one explanation.

Longer exposure to government bureaucracy is associated with lower degrees of report accuracy. One explanation is the spill-over effect of the corruption prevalent in Ugandan government offices during the time. Barr and Serra's (2010) experimental study suggests that individuals coming from a corrupt environment may also act corruptly in a new environment. Since the Ugandan public sector in early 2000s was plagued with corruption (Deininger and Mpuga, 2005), NGO managers switching from government jobs may carry the ethos to their new positions. If we interpret longer government service as having a deeper political connection with the authority, the result supports the similar finding by Chaney et al. (2011) that politically connected firms have poorer quality of accounting information than non-connected firms. Connected organisations can afford to disclose lower quality accounting information as the political connection might reduce the need to respond to regulatory pressures from the authority and the donation market.

Regarding reporting fatigue, more burdensome reporting requests from donors are significantly associated with lower levels of report 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 A2 suggests 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). One counterargument for this explanation is that donors simply adjust their demands towards organisations that are more likely to behave in a dubious manner. Since such information is unobservable, the organisation heterogeneity biases 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 organisations available to fund so that the donors had to compromise and use dubious NGOs. Second, even after controlling for the unobserved heterogeneity that relates to the organisation's ability and potential manipulation (see Section 6.3), the negative association remains (see Table A4).

6. Robustness checks

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

6.1 Alternative measures of reporting accuracy

The positive relationship between performance and reporting accuracy is not sensitive to alternative measures of deviation from Benford's Law. Appendix C presents three "critical-value" based proxies widely used in previous studies (for example Michalski and Stolz, 2013 for the Pearson's Chi-square test; and Morrow, 2014, Amiran et al., 2015 for Cho-Gaines' (2007) *d*-statistics and the Kolmogorov – Smirnov test for fit of distributions). Although these methods require a priori choice of critical value, which is prone to subjectivity, they offer ease of use and practical interpretations (see Nigrini and Mittermaier, 1997 and Brabesi et al., 2017 for discussions). Table A3 reports the results for each of the alternative test statistics. The results do not alter the findings presented in Table 1.

6.2 Endogeneity between performance and reporting accuracy

The analysis attempts to account for endogeneity, caused by endogenous sample selection, by a Heckman correction model. Yet, there are other biases that could be of concern.[24] For example, altruistic NGOs could self-commit to deliver both higher performance and increased transparency and accountability. Failing to control for the bias would diminish our claim that better performance motivates an NGO to be more transparent and accountable. Although we aim to capture these concerns with a set of control variables suggested in the for-profit literature, the concern remains possible. We propose an instrumental variable strategy to reduce the bias. Computationally, we use procedures for

[24] In Table 1, a negative correlation between the error terms from the two stages (Lambda = -2.522) implies a potential negative sample selection bias: NGOs with incomplete disclosure would have a higher potential reporting accuracy. This interpretation is contradictory to the discretion-based hypothesis. One justification is that the reservation level of reporting accuracy is heterogeneous across organisations, perhaps due to unobserved management styles. Some NGOs have the reservation level so low that they chose complete disclosure even when their reports were moderately accurate; while other NGOs whose reservation levels are significantly higher, opted for incomplete disclosing. The heterogeneity leads to the potential report accuracy of incompletely disclosing NGOs being higher than that of NGOs with complete disclosure. Once we account for this omitted heterogeneity in Table 3, the correlation becomes positive (Lambda = 3.217) as predicted.

sample selection and censored Tobit estimators with endogenous regressors proposed by Smith and Blundell (1994), Wooldridge (2010), and Semykina and Wooldridge (2010) (see Online Appendix C).

A valid set of instrumental variables, Z_i, should have a strong correlation with the evaluated performance but be independent of any strategic behaviour 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 in the error terms. Second, community characteristics (e.g. available infrastructure, prosperity indicators, 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 proportion of the group older than 55 (%) and (ii) the proportion of the group who have a connection with the concerned NGO, e.g. a staff or an NGO member (%).[25]

Our identification assumption is that certain demographics of the focus group (age and connection with the NGO) are strongly correlated with the evaluated performance of an NGO. In particular, groups with more respondents with a connection to an NGO may report higher satisfaction simply because they have had interactions with and/or benefited from the NGO and its services. In contrast, senior-age members might be limited in the choice of support activities. As summarised in a general report of this survey (Barr et al., 2003), no NGOs reported any services or activities that specifically targeted the elderly. We expect that the needs of this demographic group are less well served by the organisations, resulting in lower satisfaction scores. F-tests and coefficients in Panel B in Table 3 confirm our conjecture. Groups with more participants having a connection with the evaluated NGO tend to give higher scores; whereas groups with more senior-age members give unfavourable assessments – significant at the 5 and 1 percent level, respectively. Using Sanderson-Windmeijer multivariate F test and under-identification test with Kleibergen-Paap LM statistics, we reject our instrumentation is weak and under-identified.

We believe the instruments are exogenous to the accuracy of the financial report to a third party for three reasons. First, there is no reason why demographics of the community like age could affect

[25] 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.

the tendency of an NGO to report accurately and responsibly to a third party. The demographics could affect how transparent the NGO is with the village, but hardly with a third party (here the enumerator) coming for research purposes. 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 A5 shows the instruments do not exhibit any strong 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 3 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 3 provides the results using the Cragg's and Heckman model with endogenous regressors and the proposed instruments for $Performance_i$. We again observe a positive relation between performance and reporting accuracy. 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-statistic reported by Sanderson-Windmeijer tests (4.49), we additionally report Anderson-Rubin confidence intervals in Table 3 (see Cameron and Miller, 2015 for recommendations). The confidence interval is [0.6, 60] and above the zero point, further supporting the significantly positive relationship found in Panel A. We interpret the results, with caution, that better performing NGOs do report more accurately. With the IV strategy, we rule out the confounding effect of any unobserved motivations or reverse causality.

Table 3. 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	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 inference test: (Prob>F)	2.48	(0.095)

Notes: *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (15) in parentheses unless stated otherwise (bootstrapped standard errors with 100 replications are largely identical). The set of instrument variables is not statistically significant in the selection equation. 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 (see Cameron and Miller, 2015). 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-identificaion and under-identification tests hypothesise if the instrumentation is identified and under-identified.

6.3 Robustness to specification form

Since Heckman and Cragg's estimates rely on assumptions of normality and heteroscedasticity in the main equation error terms, we report three simple tests of normality and heteroscedasticity.[26] Figure A1

[26] Throughout the analysis, we estimate the Cragg's (1971) normal truncated model, which assumes a normal distribution of the report accuracy variable (R_i). It is, however, unsatisfactory to have the possibility that a rational NGO might plan to have a negative degree of report accuracy then be forced by other factors to provide incomplete disclosure. We experiment with the Cragg's (1971) log-normal truncated model, in which we instead take the logarithm of the report accuracy variable ($\log R_i^*$) as the dependent variable. The results are qualitatively unchanged. However, the respective heteroscedasticity test strongly rejects the null of homoscedasticity. For that reason, we estimate the Cragg's (1971) normal truncated model in the main analysis, noting the unsatisfactory possibility of a (potential) negative report accuracy.

in Appendix D reports a clear graphical resemblance between the predicted residual density and a normal distribution. In Table 1, Jaque-Bera test statistics strongly supports residual normality, while we barely fail to reject the null hypothesis of homoscedasticity at 5%. In this exercise, we relax the assumptions on functional forms and distribution types of the error terms, showing that our main result for the positive relationship remains similar.

We estimate two simple nonparametric kernel regressions (local polynomial smooth and lowess smother) in Figure A2 for a bivariate relationship between performance and reporting accuracy.[27] The nonparametric becomes complex when adding more variables into a kernel regression because it introduces locally sparse noises ("curse of dimensionality").[28] 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 A3 support the result.

Second, we perform a Robinson (1988) semiparametric estimator for sample selection model and Powell (1984) censored least absolute deviations (CLAD) for the censored Tobit model (see Online Appendix D for details). The estimators are robust to heteroscedasticity, consistent, and asymptotically normal for a wide class of error distributions. Figure A4 reports Robinson (1983) estimations for scenarios when $Performance_i$ is considered exogenous and endogenous. Using Powell's (1984) estimator, the estimate, standard error bootstrapped and clustered at district level, and bias-corrected for $Performance_i$ are 5.381, 3.318 and [4.616, 16.517] respectively. The results strongly suggest 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.

[27] 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.
[28] Das et al., 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 (2003) provides a review on nonparametric and semiparametric models dealing with endogeneity.

7. Concluding remarks

In this paper, we demonstrate the feasibility and usefulness of Benford's Law as a cost-effective way to study irregularities in financial accounts 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 also find that the underlying mechanism for nondisclosure of financial accounts and the provision of false information to be uncorrelated, suggesting that shortage of skills and resources contributed to non-disclosure. Given the fiscal constraints of government-funded regulators in developing countries, it may be necessary to mobilise the donor community to contribute towards funding such oversight mechanisms. 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. However, we find excessive reporting requirements is correlated with lower levels of accuracy. If a reasonable reporting burden is exceeded, cynicism may set it, eroding the commitment to accurate and transparent reporting.

Given the alignment of community satisfaction with accurate financial reporting, and the increased emphasis on community responsiveness and community assessments, this work provides support for the prioritisation 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 organisation's main charitable activities, we provide evidence that collecting assessments from the beneficiary communities may be more efficient.

Although our study may suffer from small sample bias, the results are robust across a range of modelling approaches: under 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.

Appendices

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:

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

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 1 records the full theoretical distribution specified by Benford's Law: 1 will appear as the leading digit 30.1% of the time, two will appear 17.6% of the time, and so forth.

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

B. Set of other explanatory variables

We describe characteristics identified in the empirical literature as possible explanatory variables of information disclosure and report accuracy (or earning managements as in the for-profit studies). Table A2 provides some descriptive statistics.

a) Standard control variables

Monitoring body: The for-profit literature posits a significant association between information transparency and the oversight board structure. While the presence of a board of trustees/directors or a governing body may enforce organisation accountability (see Amstrong et al., 2014 and citations within), an "unintended chilling effect" could instead make management less willing to voluntarily disclose information that is deemed unnecessary (Cormier et al., 2005). We include two variables that capture this idea: (i) whether the NGO has a board of directors or trustees to oversee its activities (*Board* = 1 if Yes, = 0 otherwise), (ii) whether the NGO has officially registered as a company (*Registered as Company* = 1 if Yes, = 0 otherwise).

Financing: We conjecture that NGOs that ever received a grant (*Received Grant* = 1 if Yes, 0 otherwise) would be more likely to have prepared complete records of their financial situation. The effect of receiving a grant on report accuracy, however, is unclear. An organisation may strive to provide

accurate information to avoid reputation loss in case any irregularities are exposed. A competing hypothesis is that organisations might alter financial figures to mislead potential donors about the underlying financial performance.

Reputation loss: One potential cost is the reputation loss for more established organisations when they fail to provide financial records if requested. We use *Age* (the years of existence to 2002) and Age^2 to proxy for the NGOs standing. The squared term is popular for a U-shaped relationship between the organisation's reputation and the need for voluntary disclosure. Another source of reputation concern is the tenure of the NGO manager. Fudenberg and Tirole (1995) suggest that future career concerns may lead managers to misrepresent the organisation's performance, particularly in the earlier stages of their tenure. We include the number of years that the manager has been with the NGO (*Tenure*) to examine the effect of career concerns.

Information gathering costs: Most disclosure studies control for firm size (see Bayer et al., 2010) assuming there are economies of scale in costs of information-gathering. Although for-profit studies suggest using total assets as a proxy for firm size, the information is not available for Ugandan NGOs who provide *incomplete disclosure*. Another possible drawback of using total assets as a measure in the non-profit literature is measurement error (Hofmann and McSwain, 2013). Instead, we use two variables: the number of clerical staff working for the NGOs (*Clerical staff)* and the proportion of paid employees having a tertiary education or a degree (% *Degrees*). One conjecture is that more clerical staff and more employees with higher levels of education would reduce the cost of producing information, facilitating the propensity to keep a full, and possibly more accurate, record. Staff with degrees might also act as an internal monitoring body that puts pressure on releasing complete records of the organisation's financial activities.

Reporting fatigue: Following Burger and Owens (2010), we use the number of reports per year required by granting bodies (*Reports per year*) to capture the heavier administrative costs. We expect a negative relationship between the reporting requirements and the sequential degree of accuracy. This hypothesis is consistent with Ebrahim's (2003) observation that Indian grass-root organisations are skilled at subtly resisting donor demands but exhibit sufficient compliance to maintain funding.

b) Control variables specific to the Ugandan NGO sector

Religious adherence: a binary variable of whether the NGO is affiliated with a religion (*Religion* = 1 if yes, = 0 otherwise). One mechanism whereby religion influences transparency choices is that these organisations may be more prone to (religious) moral codes of conduct (Mazar et al., 2008; Dyreng et al., 2012; McGuire et al., 2011).

Attitude towards governance. We propose four variables to capture the organisation's attitude towards internal constraints and monitoring activities. First, we use two binary variables of whether the NGO cites *Lack of funding* (= 1 if Yes, = 0 otherwise) and *Lack of skilled staff* (= 1 if Yes, = 0 otherwise) as a constraint preventing them from doing an even better job. These constraints may work as either internal hurdles towards accountability or a means for the organisation to request further funding to cover overhead expenditures. 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). Following Burger and Owens (2010), we 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 potential spill-over effects of the corruption in Ugandan government and public service departments in the 2000s (see Deininger and Mpuga, 2005 for the survey; Barr and Serra, 2010 for spill-over effects of corruption). The variable may also represent a stronger political connection of the NGO manager which may translate into less monitoring pressure from regulating bodies (Stigler, 1971; Chaney et al., 2011).

Table A2. 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 (community satisfaction)	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 related or expenses related information or none. *Complete* represents 77 NGOs who provide all revenues related and expenses related financial figures as requested. Categorisation of conformity is based on Pearson's Chi-square tests between the observed distribution and the theoretical distribution (Deviate if p-values of Chi-square are less than 5%, Conform otherwise). *Source:* Authors' analysis based on the 2002 Ugandan NGO survey data.

C. 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 Chi-square tests at the significance of 10% (*Conform* = 1 if Yes, 0 otherwise):

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

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

$$Conform = \begin{cases} 1 & \text{if } \Pr(X^2 \leq X_N^2(\alpha)) \geq 0.05 \\ 0 & \text{if } \Pr(X^2 \leq X_N^2(\alpha)) < 0.05 \end{cases} \tag{A4}$$

where N is the total number of non-zero financial items used, X_N^2 is the critical value of the Chi-square distribution at N and test power $\alpha = 0.10$. 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. Like the measure from the MAD statistic, lower values of the indices indicate that the tested data diverge further from the Benford distribution. Table A3 show similar results for the main equation.

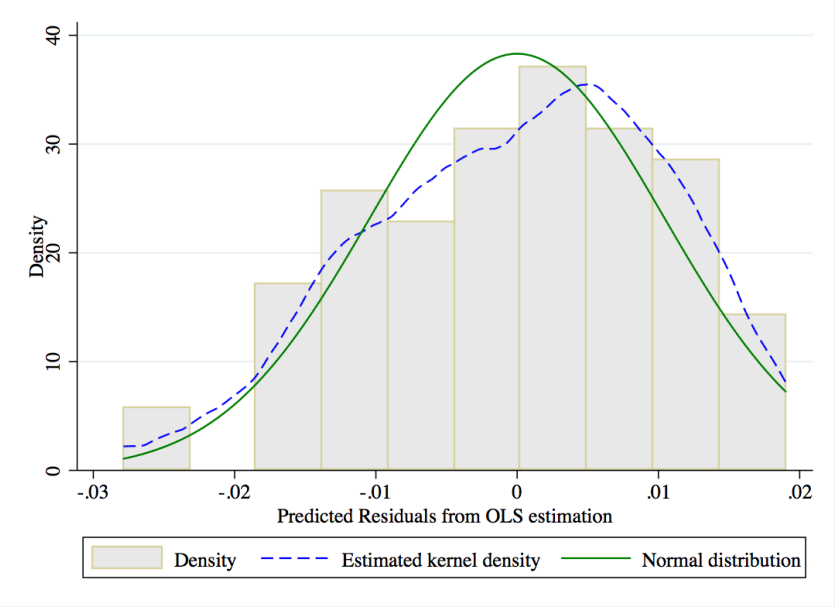
Table A3. Estimations of outcome equations for alternative measures of reporting accuracy

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 year	-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

Notes: *** 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.

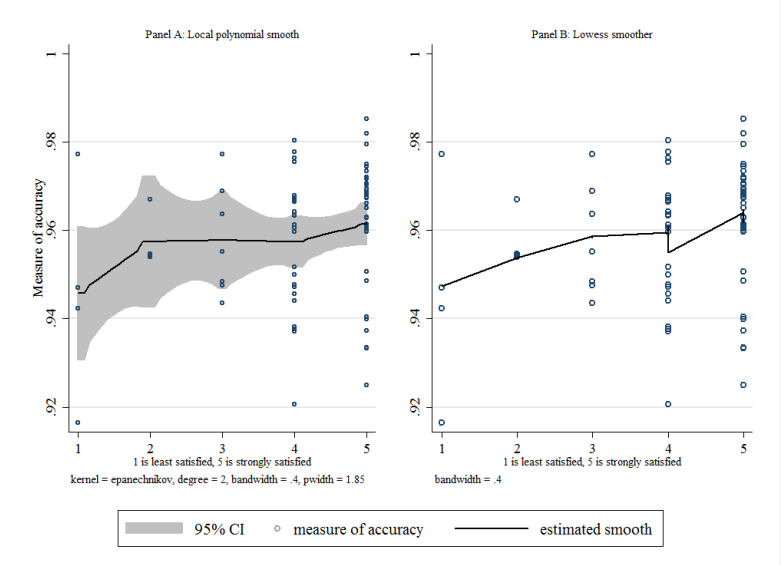
D. Non-parametric and semiparametric estimations

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



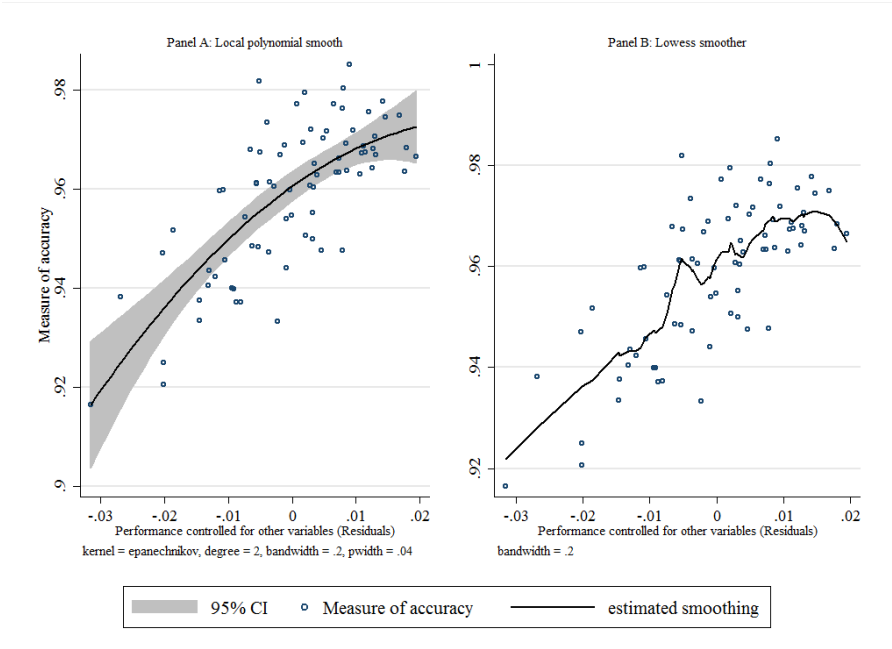
Notes: Estimated kernel density is also plotted. A clear resemblance between the kernel density and the normal distribution suggests the normality condition in residual terms is graphically valid.

Figure A2. Kernel-weighted regressions of reporting accuracy (*y)* and NGO performance (*x)*



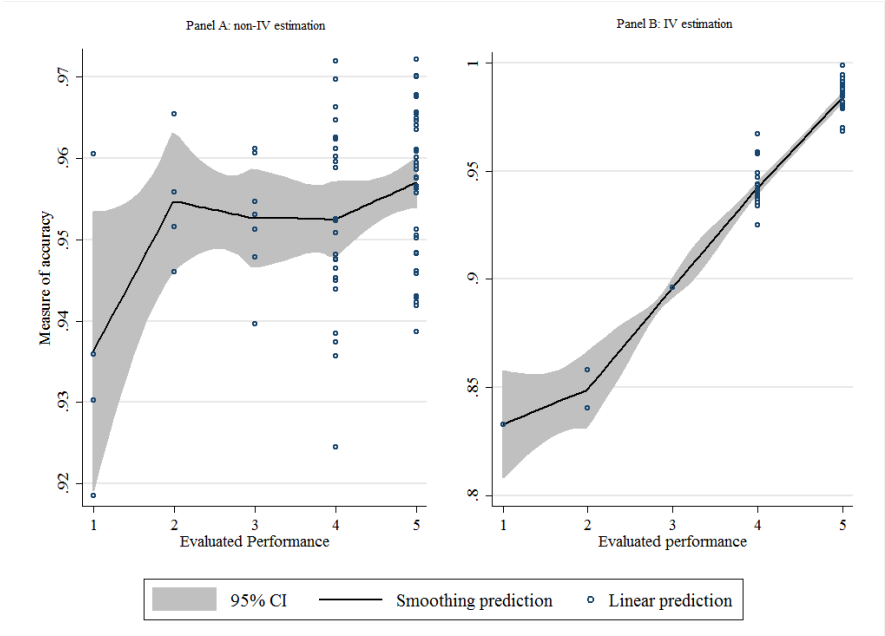
Notes: Panel A reports a kernel-weighted local polynomial regression (multiple piecewise linear estimations *locally* in a neighbourhood of *x* within a given bandwidth). Polynomials of *x* are included to improve the fit of the estimation. The default Kernel (Epanechikov) distribution and 95% confidence 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 A3. Kernel-weighted regressions of accuracy measure and relative performance (residuals)



Notes: 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 A4. Semiparametric non-IV and IV estimation following Robison (1988) selection model.



Notes: 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. Additional Tables for the IV estimations

Table A4. IV estimations of selection and outcome equations for control variables

VARIABLES	Selection equation	Outcome equation: R_i	
	$C_i(.)$	IV-Heckit	IV-Cragg
Board	37.765	-0.455	-0.586
	(104.395)	(0.674)	(0.376)
Registered as company	1,604.467**	8.659	0.112
	(723.274)	(7.205)	(3.241)
Received grant (Yes = 1)	56.767	12.245***	10.638***
	(388.568)	(3.077)	(2.374)
Age	982.209**	6.593	1.734
	(423.353)	(5.443)	(4.049)
Age^2	-719.620	-0.994	0.679
	(521.026)	(3.709)	(2.423)
Tenure length	236.746	-0.068	-0.743
	(402.950)	(3.935)	(3.045)
% Professional degrees	-843.225**	-9.860*	-7.042
	(403.710)	(5.530)	(4.626)
Clerical staff	0.313	-0.533**	-0.414*
	(22.731)	(0.222)	(0.224)
Reports requested per year	-135.144**	-0.697	-0.333*
	(62.618)	(0.546)	(0.195)
Religious Affiliation	95.982	0.231	-0.024
	(60.820)	(0.259)	(0.117)
Lack of skilled staff	208.572*	2.900**	1.502***
	(107.683)	(1.140)	(0.384)
Lack of funding	-6.471**	-0.088***	-0.043***
	(3.045)	(0.033)	(0.010)
Government as a hindrance	1,712.166***	0.885	-3.013
	(536.460)	(5.019)	(3.213)
Years working in government	-1,114.106	-2.749	0.659
	(704.692)	(6.502)	(4.930)
Member involvement	704.594*		
	(401.664)		
Number of non-zeros		0.582***	0.516***
		(0.099)	(0.083)
Lambda/sigma		3.217	10.042***
		(6.872)	(0.679)
Constant	-845.115	889.015***	899.071***
	(1,154.951)	(15.779)	(13.051)
Observations	100	100	100

Notes: *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (15) in parentheses (bootstrapped standard errors with 100 replications are largely identical). The set of instrument 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.

Table A5. 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)

Notes: 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.

References

Acht, M., Mahmoud, T. O., & Thiele, R. (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, 20–33.

Amiram, D., Bozanic, Z., & Rouen, E. (2015). Financial statement errors: evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies*, *20*(4), 1540-1593.

Barabesi, L., Cerasa, A., Cerioli, A., & Perrotta, D. (2017). Goodness-of-fit testing for the Newcomb-Benford law with application to the detection of customs fraud. *Journal of Business & Economic Statistics*, 1-13.

Barr, A., & Fafchamps, M. (2006). A client-community assessment of the NGO sector in Uganda. *The Journal of Development Studies*, 42(4), 611-639.
Barr, A., Fafchamps, M. and Owens, T. (2003), *Non-Governmental Organizations in Uganda*, CSAE, Oxford University, Oxford. (Report to the Government of Uganda).
Barr, A., Fafchamps, M., & Owens, T. (2005). The governance of non-governmental organizations in Uganda. *World Development*, 33(4), 657-679.
Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *The American Economic Review*, *96*(5), 1652-1678.
Bettin, G., Lucchetti, R., & Zazzaro, A. (2012). Endogeneity and sample selection in a model for remittances. *Journal of Development Economics*, *99*(2), 370-384.
Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of accounting and economics*, *50*(2), 296-343.
Brass, J. N. (2012). Why do NGOs go where they go? Evidence from Kenya. *World Development*, *40*(2), 387-401.
Burns, N., & Kedia, S. (2006). The impact of performance-based compensation on misreporting. *Journal of financial economics*, *79*(1), 35-67.
Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317-372.
Chen, Q. (2016). Director Monitoring of Expense Misreporting in Nonprofit Organisations: The Effects of Expense Disclosure Transparency, Donor Evaluation Focus and Organization Performance. *Contemporary Accounting Research*. http://dx.doi.org/10.1111/1911-3846.12218.
Chaney, P. K., Faccio, M., & Parsley, D. (2011). The quality of accounting information in politically connected firms. *Journal of Accounting and Economics*, *51*(1), 58-76.
Cho, W.K.T. and Gaines, B.J. (2007) Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance. *The American Statistician*. 61, 218–223.
Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, organizations and society*, *33*(4), 303-327.
Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 1431-1451.
Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, 829-844.
Deininger, K., & Mpuga, P. (2005). Does greater accountability improve the quality of public service delivery? Evidence from Uganda. *World development*, *33*(1), 171-191.
Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, *50*(2), 344-401.
Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of forensic accounting*, *5*(1), 17-34.
Dye, R. A. (1986). Proprietary and Non-proprietary Disclosures. *The Journal of Business*, *59*(2), 331-66.
Dyreng, S. D., Mayew, W. J., & Williams, C. D. (2012). Religious social norms and corporate financial reporting. *Journal of Business Finance & Accounting*, *39*(7-8), 845-875.
Ebrahim, A. (2003). Accountability in practice: Mechanisms for NGOs. *World Development*, 31(5), 813-829.
Edwards, M., & Hulme, D. (Eds.). (1996). *Beyond the magic bullet: NGO performance and accountability in the post-cold war world* (p. 1). West Hartford, CT: Kumarian Press.
Fruttero, A., & Gauri, V. (2005). The Strategic Choices of NGOs: Location Decisions in Rural Bangladesh1. *Journal of Development Studies*, *41*(5), 759-787.
Fang, H., & Gong, Q. (2017). Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked. *The American Economic Review*, *107*(2), 562-591.
Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, *3*(4), 305-360.
Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of accounting research*, 193-228. Fang, H., & Gong, Q. (2017). Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked. *The American Economic Review*, *107*(2), 562-591.

Judge, G., & Schechter, L. (2009). Detecting problems in survey data using Benford's Law. *Journal of Human Resources*, *44*(1), 1-24.
Healy, P. M. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics*, *7*(1), 85-107.
Herzlinger, R. E. (1996). Can public trust in nonprofits and governments be restored?. *Harvard Business Review*, *74*(2), 97-107.
Hill, T.P (1995). The first digital phenomenon. *American Scientist.* 86(4):368-363.
Hofmann, M. A., & McSwain, D. (2013). Financial disclosure management in the non-profit sector: A framework for past and future research. *Journal of accounting literature*, *32*(1), 61-87.
Keating, E. K., & Frumkin, P. (2003). Reengineering nonprofit financial accountability: Toward a more reliable foundation for regulation. *Public Administration Review*, *63*(1), 3-15.
Kothari, S. P., Shu, S., & Wysocki, P. D. (2009). Do managers withhold bad news?. *Journal of Accounting Research*, *47*(1), 241-276.
Lang, M., & Lundholm, R. (1993). Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of accounting research*, 246-271.
Lee, C. W. J., Li, L. Y., & Yue, H. (2006). Performance, growth and earnings management. *Review of Accounting Studies*, *11*(2-3), 305-334.
Leuz, C., & Wysocki, P. D. (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research*, *54*(2), 525-622.
Li, S., Peitz, M., & Zhao, X. (2016). Information disclosure and consumer awareness. *Journal of Economic Behavior & Organization*, 128, 209-230.
McGuire, S. T., Omer, T. C., & Sharp, N. Y. (2011). The impact of religion on financial reporting irregularities. *The Accounting Review*, *87*(2), 645-673.
Michalski, T., & Stoltz, G. (2013). Do countries falsify economic data strategically? Some evidence that they might. *Review of Economics and Statistics*, 95(2), 591-616.
Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 380-391.
Miller, G. S. (2002). Earnings performance and discretionary disclosure. *Journal of accounting research*, *40*(1), 173-204.
Morrow, J. (2014). Benford's Law, Families of Distributions and a Test Basis. *CEP Discussion Paper No 1291*.
Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy*, *115*(2), 200-249.
Silva, J., Santos, M. C., Tenreyro, S., & Windmeijer, F. (2015). Testing competing models for non-negative data with many zeros. *Journal of Econometric Methods*, *4*(1), 29-46.
Stigler, G. J. (1971). The theory of economic regulation. *The Bell journal of Economics and Management Science*, 3-21.
Teoh, S. H., & Hwang, C. Y. (1991). Nondisclosure and adverse disclosure as signals of firm value. *Review of Financial Studies*, *4*(2), 283-313.
Thakor, A. V. (2015). Strategic information disclosure when there is fundamental disagreement. *Journal of Financial Intermediation*, *24*(2), 131-153.
Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.
Woodwell, H.W., & Bartczak, L. (2008). *Is Grantmaking Getting Smart? A National Study of Philanthropic Practice.* Washington, D.C.: Grantmakers for Effective Organisations.
Zitzewitz, E. (2012). Forensic economics. *Journal of Economic Literature*, 50(3), 731-769.

Appendices 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 and Trudy Owens (University of Nottingham)

****

A. Conceptualising 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 organisation. This proposition is consistent with the finding that costly accounting procedures could harm the monitoring process as the organisation now faces a higher trade-off between upward accountability and downward accountability. Once the organisation'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 maximisation problem:[1]

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

[1] The illustration is inspired by the selection mechanism in Bettin et al. (2012) on remittances.

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.[2] 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 reports 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 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 $R_i > 0$.[3] 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

[2] 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.

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

last assumption is to imply the NGO's preference on the consumption set of performance measures (such as beneficiary's feedback) is convex.[4]

- $U_P^i = \frac{\partial U_i}{\partial P_i} > 0,\ \ U_{PP}^i > 0,\ U_R^i = \frac{\partial U_i}{\partial R_i} \geq 0,\ \ U_{RR}^i > 0$. 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.

Proposition 1. If $\tau > 0$, there always exists a unique solution $(P_i, R_i) \in \mathbb{R}_{\geq 0}$ to (1).[5] 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 $U_R^i = 0$, 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 $(E_i, 0)$.[6] In this case, the construction of U_R^i governs both mechanisms underlying disclosure and accuracy decisions.

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

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

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

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 maximised utility. Denote $G(P_i) = E_i - \tau C(.) - P_i$ the constraint set of the main problem. As

[4] 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.

[5] 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.

[6] 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 $U_P > 0$.

$\tau > 0$, $\{P_i: G(P_i) \geq 0, x \geq 0\}$ 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 $t \in (0,1)$:

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

Thus, $tx_1 + (1-t)x_1'$ is a feasible point. Since $U_R^i > 0$ and $U_i(P_i, .)$ is strictly quasi-concave on P_i, we have that:

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

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 $u_1 = \max_{E_i - \tau > P_i \geq 0} U_i(P_i, R_i) < U_i(E_i, 0)$, 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 $\max_{E_i - \tau > P_i \geq 0} U_i(P_i, R_i) < U_i(E_i, 0)$, $\tau > 0$ and $U_R^i > 0$, we have $\lim_{R_i \to 0} U_i(P_i, R_i) < U_i(E_i, 0)$. 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 $u_1 = \max_{E_i - \tau > P_i \geq 0} U_i(P_i, R_i) < U_i(E_i, 0)$, 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 maximises 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. Given the existence, if incomplete disclosure occurs, $\mathrm{R^*} = 0$, the lemma is bounded.

If complete disclosure occurs, consider the main maximisation 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 $\mathrm{R_i}$: $\max_{0 \leq \mathrm{P_i} \leq \mathrm{E_i} - \tau} \mathrm{U_i}(\mathrm{E_i} - \tau - \mathrm{R_i}, \mathrm{R_i})$.

Since we only have one choice variable, supermodularity 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:

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

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 minimise KLIC is equivalent to maximising the expected log-likelihood $E(L^*)$ and derives the following likelihood statistics adapted in our context:

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

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. $\hat{\omega}_n^2$ is the estimated variance of the pointwise log-likelihood ratio calculated as:

$$\hat{\omega}_n^2 \equiv \frac{1}{n}\sum_1^n \left[\ln\frac{f(\hat{\beta}_n)}{f(\hat{\theta}_n)}\right]^2 - \left[\frac{1}{n}\sum_1^n \ln\frac{f(\hat{\beta}_n)}{f(\hat{\theta}_n)}\right]^2 \quad \text{(OA.8)}$$

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.

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:

$$R_i^* = \alpha_1 + \gamma x + X_i'\beta_1 + u_i \quad \text{(OA.9)}$$

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

Some studies have addressed the simultaneity problems of heterogeneity and selectivity (see Angrist, 2001 and 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, Semykia 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: $\mathrm{C_i}(.) = \mathrm{Z_i'}\gamma_1 + \mathrm{S_i'}\alpha_1 + \mathrm{u_{1i}}$. Obtain the estimated inverse Mills ratios: $\hat{\lambda}_{i2}$. Second, estimate the adjusted main equation $\mathrm{R_i^*} = \mathrm{X_{1i}'}\beta_2 + \beta_{\mathrm{IV-Heckman}}\mathrm{x} + \theta_1\hat{\lambda}_{i2} + \mathrm{v_{i2}}$ by LIML using instruments $(\mathrm{Z_i}, \hat{\lambda}_{i2})$ 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: $R_i^* = X_{1i}'\beta_3 + \beta_{\text{IV}-\text{Cragg}}x + \theta_2\hat{v} + v_{2i}$. The estimates are consistent and the standard errors are clustered at district level and bootstrapped with 100 replications.

The estimates $\beta_{\text{IV}-\text{Heckman}}$ and $\beta_{\text{IV}-\text{Cragg}}$ 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

To illustrate the Robinson (1988) estimator, we rewrite the sample selection model as follows:

$$\mathrm{R_i} = \mathrm{m(x_i)} + \mathrm{X_i'\beta} + \lambda(\mathrm{S_i}) + \varepsilon_\mathrm{i} \quad \text{(OA.9)}$$

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 component of the equation, consisting of other covariates for the outcome and selection equation. $\lambda(S_i)$ is the inverse Mills ration obtained from the selection regression. The double residual estimator of Robinson (1988) is obtained by:

$$\underbrace{R_i - E[R_i|x_i]}_{u_1} = (X_i - E[X_i|x_i])'\beta + \lambda(S_i) - E[\lambda(S_i)|x_i] + \varepsilon_i = \underbrace{(T_i - E[T|x_i])\gamma}_{u_2} + \varepsilon_i \quad \text{(OA.10)}$$

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 the clustered error terms to account for the possibility that $\hat{\lambda}(S_i)$ does not converge to its parametric estimation.[7] We also experiment with different trimming levels incrementally from 0.00 to 0.05. Since the results are generally similar, we report the default trimming level of 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:

$$R_i = m(x_i) + X_i'\beta + \lambda(S_i) + \rho\tau_i + \eta \quad \text{(OA.11)}$$

The partially linear model can be estimated by conditioning on $\mathrm{x_i}$:

[7] 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).

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

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. $\hat{v} = x - Z'\hat{\pi}$. Again, to account for the residuals being estimated, we bootstrap the clustered error terms at 50 replications.

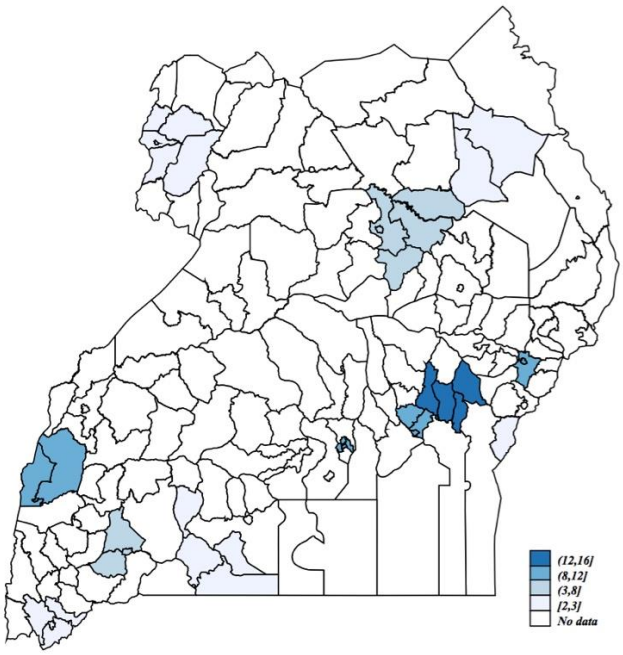
For a semiparametric censored Tobit model of $R_i = \max\{0, X_i'\alpha + \varepsilon\}$, Powell (1984) proposes estimation of the unknown parameters α by the minimiser $\hat{\alpha}_{CLAD}$ (censored least absolute deviations):

$$\hat{\alpha}_{CLAD} = \text{argmin}\, Q_n(\alpha) = \text{argmin}\, \frac{1}{n}\sum_{i=1}^{n}[R_i - \max\{0, X_i'\alpha + \varepsilon\}] \quad \text{(OA.12)}$$

Bunchinsky (1998) provides an iterative linear programming algorithm (IPLA) to computationally estimate the parameters.

E. The geography of the surveyed NGOs in 2002

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



Notes: Darker blue represents more NGOs (the number is in parenthesis) were drawn from the district

References for Online Appendices

Almond, D., & Xia, X. (2017). Do nonprofits manipulate investment returns?. *Economics Letters*, *155*, 62-66.

Angrist, J. D. (2001). Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice. *Journal of Business & Economic Statistics*, 2-16.

Bettin, G., Lucchetti, R., & Zazzaro, A. (2012). Endogeneity and sample selection in a model for remittances. *Journal of Development Economics*, *99*(2), 370-384.

Blundell, R., & Smith, R. J. (1994). Coherency and estimation in simultaneous models with censored or qualitative dependent variables. *Journal of Econometrics*, *64*(1), 355-373.

Buchinsky, M. (1994). Changes in the US wage structure 1963-1987: Application of quantile regression. *Econometrica: Journal of the Econometric Society*, 405-458.

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics*, *36*(3), 231-250.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, *25*(3), 303-325.

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of econometrics*, *39*(3), 347-366.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 931-954.

Sanderson, E., & Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, *190*(2), 212-221.

Semykina, A., & Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, *157*(2), 375-380.

Smith, R. J., & Blundell, R. W. (1986). An exogeneity test for a simultaneous equation Tobit model with an application to labor supply. *Econometrica: Journal of the Econometric Society*, 679-685.

Schwiebert, J. (2015). Estimation and interpretation of a Heckman selection model with endogenous covariates. *Empirical Economics*, *49*(2), 675-703.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.