

University of Nottingham
UK | CHINA | MALAYSIA

Donors and Recipients: Charities, NGOs and Aid

Canh Thien Dang

Centre for Research in Economic Development and International Trade (CREDIT)

School of Economics, The University of Nottingham

The thesis is submitted to The University of Nottingham for the degree of Doctor of Philosophy in Economics

September 2018

Abstract

Non-profit organisations (NPOs) play an integral part in delivering public services and channelling development aid around the world. This thesis contains three papers on behaviour, motivations, and the accountability of these organisations.

The first paper studies charitable spending and financial transparency of public charities in a developed NPO sector. Its accountability has attracted considerable scrutiny after high-profile scandals related to misuses of funding and donations. We use Benford's Law to evaluate the potential errors in public financial figures of approximately 10,000 UK charities. We estimate 25% of the sample misreport their financial information. We show unless spending on governing activities (commonly referred to as overhead cost) is sufficiently high, NPOs with higher spending on charitable activities (commonly referred to as programme ratio) will be more likely to misreport their financial information. We suggest that tighter monitoring might be ineffective to increase the sectoral transparency and accountability. Our analysis also challenges the common practice in the sector that programme ratios and overhead costs should be used as reliable indicators for non-profit accountability.

The second paper focuses on performance evaluation in a development NGO context. Little has been done to understand the connection between the effectiveness of NGOs and their financial accountability. Using Benford's Law and a representative survey of the Ugandan NGO sector, we also estimate 25% of the NGOs might provide misleading accounts. We observe NGOs with better ratings from their beneficiaries are more likely to submit credible financial data. The result contradicts the belief that upward accountability demands crowd out serving the client community. We find no evidence for a strong correlation between the decision to withhold requested financial information and the decision to report inaccurately, with the former potentially due to limited capacity and skills. The results suggest a more significant role for beneficiary assessments in monitoring the sector rather than increasing demand for financial reporting.

The third paper studies behavioural motivators of development NGOs. We identify two motivations underlying NGO decision to diversify activities, namely to reduce idiosyncratic risk to funding and to derive personal gains. We exploit a historic flood in mid-2007 in Uganda and variations in contractual funding toward affected areas to show that NGOs diversify less in response to increased income from contractual funding. This result is consistent with a theoretical prediction that the decision is driven by the need to reduce risk rather than to capture private gains. Donors can help development NGOs focus on their primary mission by providing a commitment to funding.

Acknowledgement

I am grateful to many people without whom the completion of my PhD would not have been possible. First, I am deeply thankful to my supervisors Oliver Morrissey and Trudy Owens for their continuous support, encouragements, and invaluable advice. I thank them for showing me their approaches to development economics and their views on various strands of empirical studies, which have helped shape my understanding of the field. They have also offered guide and support on how to navigate the PhD, writing in economics, and life beyond the PhD.

The thesis has significantly benefited from numerous discussions and comments from academic faculty, fellow PhD students, academic visitors at Nottingham, and conference participants at different places. I especially thank Richard Upward and Frank Windmeijer for their thorough suggestions that have improved the papers. For the opportunity to pursue a PhD and to lengthen the list of attended conferences, I wish to acknowledge generous financial support from the Economics and Social Research Council (grant number ES/JS500100/1) and the Nottingham School of Economics. I am thankful to the administrative staff at Nottingham for helping me with the paperwork and making my life as an overseas PhD student more manageable. I particularly thank Sarah Nolan, Lydia Topliss, Hayley McCalla, and Frances Twiddy for many joyful conversations. For a short research visit at the University of Brescia (Italy), I am grateful to Rafaelle Miniaci and the financial support from the IAAE and ACRI for their hospitality.

For making the PhD experience so much more enjoyable, I am thanking Hayat, Derrick, Mohammed, Pryanka, Bernardo, Thomas, Juergen, Francesca, Roel, Apoorva, Vaselis, Valeria, Kyung, Martina, Lionel, Douglas, Marcy, Arno, Basile, Diego, Bjorn, Oanh and all my colleagues and friends at Nottingham. I thank "House 34 Mayfield" and anh Thanh for introducing me to a number of new non-academic activities and rekindling my hobby in photography. Before I came to Nottingham, Thomas Holden has mentored me at Surrey and inspired me to do a PhD in Economics. His passion for research and technical rigidity are an example for me to strive for.

However, I owe everything to my parents who have believed in me and supported me all the way in this academic journey. My parents and my younger sister have been with me every step, providing life advice and unconditional love to help me finish the journey. The four years at Nottingham would never be as complete without the companionship and joy and love I share with Trang. We have come so much far together, and I am thankful to her for being a part of my life. Not only that, she directly contributed to the writing of this thesis by patiently teaching me how to put up a neat table of contents and the lists of figures as evidenced below.

Table of Contents

Abstract **1**

Acknowledgement **2**

Chapter 1. Introduction **7**

Chapter 2. Transparency or charitable services? **10**

2.1. Introduction **11**

2.2. Benford's Law and the UK Third Sector Research Data **14**

2.2.1. Forensic Economics Studies and Benford's Law 14

2.2.2. Data 17

2.2.3. Measures of the level of reporting inaccuracy 20

2.2.4. An illustration of Benford's Law 22

2.2.3. A theoretical model of optimal misreporting **24**

2.3.1. Basic building blocks 24

2.3.2. Timeline 25

2.3.3. Payoffs 26

2.3.4. Comparative statics and testable predictions 29

2.4. Econometric methodology **32**

2.4.1. Empirical specification 32

2.4.2. Empirical strategy: traditional IV and Lewbel's (2012) approach 35

2.5. Empirical results **37**

2.5.1. Descriptive statistics 37

2.5.2. Regression results 41

2.6. Sensitivity Analyses **45**

2.7. Concluding remarks **46**

2.8. Appendix For Chapter 2 ..... **48**

2.8.1. Proof of Lemma 1 50

2.8.2. Proof of Proposition 1 50

2.8.3. Proof of Proposition 2 51

2.8.4. Additional summary statistics 52

2.8.5. Lewbel' (2012) IV estimator 52

2.8.6. Robustness checks 53

2.8.7. Additional Tables 57

Chapter 3. Do better performing NGOs report more accurately? **58**

3.1. Introduction **59**

3.2. The Ugandan NGO data **62**

3.3. Using Benford's Law to detect potentially fraudulent NGOs **64**

3.4. A Conceptual Framework **68**

3.4.1. Why do NGOs provide incomplete financial information? 68

3.4.2. Do NGOs with better performance report more accurately? 70

3.5. Econometric Methodology and Analysis 72
3.5.1. Deliberate and unintended deviations from Benford's Law 72
3.5.2. The empirical specification 75
3.5.3. Descriptive statistics 76
3.5.5. Report accuracy, NGO performance and other determinants 82
3.6. Robustness checks 83
3.6.1. Alternative measures of reporting accuracy 83
3.6.2. Endogeneity between performance and reporting accuracy 83
3.6.3. Robustness to specification form 86
3.7. Concluding remarks 88
3.8. Appendix For Chapter 3 89
3.8.1. The Benford's distribution of first digits 89
3.8.2. Robustness to alternative measures of conformity to Benford's Law 90
3.8.3. Non-parametric and semi-parametric estimations 92
3.8.4. Balance tests for excludability argument of the instruments 94
3.8.5. A Sample of Financial Information Requested in the 2002 survey 95
3.8.6. Technical details 97
Chapter 4. What motivates development NGOs? 108
4.1. Introduction 109
4.2. Why do development NGOs diversify? 113
4.2.1. Diversification to reduce risks related to funding 113
4.2.2. Diversification for private gains 114
4.3. A conceptual model on incentives and diversification 116
4.3.1. Building blocks 117
4.3.2. Optimal actions and comparative statistics. 120
4.4. Empirical analysis 124
4.4.1. Data: the 2008 Ugandan NGO survey 124
4.4.2. Empirical strategy 126
4.4.3. Empirical results 134
4.4.4. Assessing the excludability assumption of the instrumentation 138
4.4.5. Alternative identification strategy 142
4.5. Concluding remarks 145
4.6. Appendix For Chapter 4 146
4.6.1. Derivation of the Certainty Equivalent (Equation 4) 146
4.6.2. Proof of Corollary 2 147
4.6.3. Additional Tables and Figures 149
4.6.4. Balance tests and additional tables 151
4.6.5. Robustness to excluding NGOs whose managers work for many NGOs 157
4.6.6. Using clustered errors with a small number of clusters and uneven sizes 157
References 163

List of Tables

Table 2.1. Probability predicted by Benford's Law for the leading digits 16
Table 2.2. Descriptive statistics of the main variables 39
Table 2.3. Descriptive statistics by conformity to the Benford distribution 40
Table 2.4. Main results for determinants of misreporting for British charities 44
Table 2.5. Diagnostic tests for the IV estimations 45
Table 2.6. Fixed effect models for misreporting over two periods 46

Table 2.A1. Financial items reported by the NPOs 48
Table 2.A2. Replications using other indices 56
Table 2.A3. Heckit estimator for missing observations for the four indices 55
Table 2.A4. Reporting inaccuracy and NPO's observable characteristics 57

Table 3.1. Descriptive statistics of Ugandan NGOs 77
Table 3.2. Estimations of the selection and outcome equations 79
Table 3.3. Model selection tests 81
Table 3.4. IV estimations of the selection and outcome equations 86

Table 3.A1. Probability predicted by Benford's Law for the leading digits 89
Table 3.A2. Estimations for alternative measures of accuracy 91
Table 3.A3. A balance test for NGO characteristic and instruments 94
Table 3.A4. Sample of Financial Information Requested in the 2002 survey 95
Table 3.A5. Correlations of closely relevant measures 106
Table 3.A6. Robustness to different measures of Performance 106
Table 3.A7. Estimates of selection and outcome equations for control variables 107

Table 4.1. Descriptive statistics for Ugandan NGOs in 2008 129
Table 4.2. First stage estimation for the instrument $FLOOD - FFECTEDi$ 133
Table 4.3. OLS and IV Estimations for diversification on incentives 136
Table 4.4. OLS and IV Estimates for diversification on incentives 137
Table 4.5. IV estimates using restricted sample: EVER_AFFECTED = 1 142
Table 4.6. Panel Estimations of diversification on incentives 145

Table 4.A1. Robustness to excluding NGOs working on Emergency Relief 149
Table 4.A2. Balance checks on working in vulnerable areas 152
Table 4.A3. Extreme floods across Ugandan districts from early 1980 to 2017 153
Table 4.A4. The number of NGOs working in each district in 2007 154
Table 4.A5. Balance checks for NGOs on revenue composition 155
Table 4.A6. Correlation between net ODA and aid and monthly rainfalls 156
Table 4.A7. Robustness to excluding NGOs whose managers work for many NGOs 157
Table 4.A8. Robustness to using pairs clustered bootstrapped procedure 158

List of Figures

Figure 2.1. The UK Third Sector Research Data and the Benford distribution 23
Figure 2.2. Timeline of the three stages.....26
Figure 2.3. Histograms of the measures of misreporting and key expenses 38
Figure 2.4. The predictive marginal effects on inaccuracy of financial reports.....42

Figure 2.A1. How sign $\partial b * \partial g$ and sign T(g) behave when g varies in $(0, +\infty)$.....51
Figure 2.A2. Measures of misreporting (MAD) by quantiles of NPO characteristics.....52
Figure 2.A3. Robustness to varying the cut-offs of non-zero financial points 53
Figure 2.A4. Robustness to excluding various ranges of outliers.....54

Figure 3.1. The Ugandan NGO data and conformity to Benford's Law.....67
Figure 3.2. Three types of NGOs regarding disclosure and report accuracy 73
Figure 3.3. Flowchart illustrates steps of our analysis.....74
Figure 3.1. Descriptive relation between report accuracy and performance 77

Figure 3.A1. Density plot of predicted residuals from OLS of the main equation 92
Figure 3.A2. Kernel-weighted regressions of accuracy (y) and NGO performance (x).....92
Figure 3.A3. Kernel-weighted regressions of accuracy and performance (residuals) 93
Figure 3.A4. Semiparametric non-IV and IV estimation following Robison (1988) 93

Map 3.A1. Districts in the 2002 Ugandan NGO survey.....104

Figure 4.1. Timeline for the model.....119
Figure 4.2. The geographical coverage of NGOs in 2007 125
Figure 4.3. The prevalence of categorised activities in 2007 126
Figure 4.4. The average rainfall (mm) in July from 1901 to 2015.....132

Figure 4.A1. Histograms of the two dependent variables 149
Figure 4.A2. Ugandan districts most affected by the 2007 flood.....150
Figure 4.A3. Ugandan districts most vulnerable to extreme floods from 1988 to 2017..150
Figure 4.A4. OLS regression of FLOOD_AFFECTED on NGOs characteristics 151

Chapter 1. Introduction

The past four decades have seen a rapid expansion of philanthropy, civil society, and non-profit organisations around the world. In developed countries, non-profits employ on average 7.5 per cent of the workforce and participate profoundly in the production and distributions of public goods and services (Salamon, 2010). In developing countries, non-governmental organisations have increasingly become the critical subcontractors of aid disbursement and the significant providers of development projects (Acht et al., 2015). Despite these promising developments, the governance and its implications for behaviour and performance of the non-profit sector remain understudied by academics and poorly understood by policymakers (Aldashev et al., 2015). Addressing this gap is essential for at least three reasons. First, the literature on this issue often cites the lack of effective measures of performance and misbehaviour as the primary constraint toward understanding the many mechanisms underlying the sector. Without effective performance measures, funding agencies and regulatory bodies are restricted regarding the ability to monitor and evaluate the organisations (Aldashev and Narrava, 2018). Second, non-profits are not immune to scandals of corruption and misuses of funding. Given the disconnection between the funders and the end users, non-profits have ample opportunities and incentives to manipulate and self-benefit at the cost of the intended beneficiaries. The voice of the beneficiaries in monitoring and assessing the sector is also often neglected in developing countries, mainly due to the excessive costs of conducting beneficiary-based evaluations of the subcontractors who carry out the development projects (Ebrahim, 2003; Mansuri and Rao, 2012). Finally, understanding the governance mechanisms and the motivations underlying pro-social behaviours of non-profits is beneficial to donors and the regulatory bodies when designing incentive packages to promote charitable activities and reducing misbehaviour of non-profits and NGOs.

The thesis tries to uncover some of the mechanisms and characteristics underlying the governance and behaviour of non-profits. There are three stand-alone, but interconnected, papers revolving around three main issues: (i) the role of information transparency and the trade-off between the cost of accountability and spending on charitable activities in a well-developed non-profit market; (ii) the connection between information transparency and the effectiveness of NGOs in a development context; and (iii) the primary motivations behind the decision-making process of development NGOs.

The first paper (Chapter 2) examines the relationship between an innovative measure of financial misreporting and two popular measures of non-profit performance, namely the programme ratio (the fraction of total income spent on charitable activities)

and the overhead cost (the fraction of total income spent on administrative and accounting activities). The chapter is of interest to policymakers, donors and charities for at least three reasons. First, we provide the first systematic analysis using an innovative and easily scalable method based on Benford's Law to evaluate the irregularities in the financial data of non-profits and charities. Second, we speak to the current demand among the public and donors for increased scrutiny of the charity sector. We find approximately 25% of a representative sample of the British charities potentially have inaccuracies in their reported financial data. Finally, we revisit a popular debate that, despite the quest for increased accountability, financial transparency and associated expenses should not come at the cost of spending on charitable missions. We cast doubt over the belief that increased overhead expenses would necessarily result in improved governance quality and accountability. Instead, we argue that without sufficient support for both charitable activities and governance spending, the pressure for transparency from the public and donors might distort a charity's operation and negatively impact their charitable agendas. We build a simple model to demonstrate that a higher level of productive effort (on charitable activities) by a non-profit does not necessarily induce less bias in a report about the organisation's value to the donor, nor does a higher level of governance spending. Empirically, we show that unless spending on governing activities is sufficiently large (around 15% of the total income), charities with higher programme ratios will have a lower level of financial reporting accuracy. The chapter calls for different inputs to evaluate the transparency and monitor activities of charities in a well-developed environment.

The second paper (Chapter 3) addresses the lack of evidence on using beneficiary-based evaluations as an improved way to assess development effort and identify misbehaviour of development NGOs. The paper also relates to the literature on regulations and NGO monitoring by local communities (Mausuri and Rao, 2012). As in Chapter 2, we use Benford's Law to detect irregularities in a representative sample of Ugandan NGOs. We also find 25% of the sample potentially misreport their financial information when requested by an external body. We extend Chapter 2 by examining the different mechanisms underlying NGO decisions to withhold requested information and the decision to report inaccurately. There is no evidence that the two decisions are related, with the former potentially due to limited capacity and skills of the organisations. We also observe a positive link between financial transparency and NGO performance. That is, NGOs with better performance ratings from their beneficiaries are more likely to submit credible financial information. The link is unlikely to be driven by unobserved heterogeneity, assumptions of functional and distributional forms, or the different ways of constructing the measures of financial transparency. Chapter 3 suggests a more significant role for

beneficiary assessments in monitoring the sector instead of increased demand for financial reporting, which Chapter 2 demonstrates to be an unreliable indicator of accountability.

The third paper (Chapter 4) slightly departs from the focus on information transparency in the first two papers. Instead of the reporting behaviour, we examine the mechanisms underlying activity diversification, another understudied behaviour of development NGOs. Understanding the mechanisms could shed light on how to design behavioural motivators for pro-social activities. Indeed, despite often commencing with a single mission, a number of development NGOs have expanded over the years and embarked on other charitable and business activities. As activity diversification could lead to management inefficiency and misallocation of human and funding resources, it remains a puzzle as to what motivates development NGOs to diversify. There are two answers often raised by academics and practitioners, namely, to reduce idiosyncratic risk to funding and to derive personal gains. We model the two motivations in a simple framework and show that a higher incentive that motivates value-creating effort will decrease (increase) diversification for NGOs that are primarily motivated by reducing risk (respectively, deriving private gains). Using the second wave of the Ugandan NGO sector survey in Chapter 3, we empirically test the predictions. We exploit a historic flood in mid-2007 and the surge in international aid towards the most affected areas to examine how Ugandan NGOs change their diversification behaviour following the changes in incomes from value-based contractual funding. We find NGOs diversify less in response to increased income from contracts. The result is consistent with the theoretical implication that the diversification decision might be driven by the need to reduce risk rather than to capture private gains. Our results call for further support from donors and funding agencies to help development NGOs focus on their main missions. Donors could design funding packages that ensure a long-term commitment to funding, enabling NGOs to work on development efforts and projects that often take time to realise.

Overall, the thesis aims to push forward the growing literature on the governance of behaviour and performance assessment in the non-profit sector. By advocating the use of innovative methods such Benford's Law and community-based evaluations, the first two papers provide the first evidence on how to best evaluate the non-profit sector in both developed and developing countries. By emphasising the importance of funding continuity in shaping the decision-making process of NGOs working in volatile areas, the third paper on activity diversification calls for stable commitments from donors and the funding agencies. Designing specific financial packages for overhead costs and long-term support could be the first step to improve the accountability of the sector and maintain the benefits that come with grassroots NGOs. Chapter 5 discusses concluding remarks and a related research agenda arising from the papers.

Chapter 2.

Transparency or charitable services?

Evidence from investigating British charities*

JEL Classifications: L31; L44; D82; H83; H49

Keywords: non-profits; misinformation; public provision of financial reports; Benford's Law; heteroscedastic-based instruments.

* The paper benefits from discussions with Ronelle Burger, Raffaele Miniaci, Alex Possajennikov, Daniel Siedmann, Richard Steinberg, Frank Windmeijer, two anonymous referees, participants at the ISTR Bi-Annual Conference (Stockholm), the RES Symposium of Junior Researchers (Bristol), and the SSES Annual Congress (Lausanne).

"Transparency is great, but not at the cost of a charity's services."

(Asheem Singh, Director at ACEVO, The Guardian, 2015)

2.1. Introduction

Misrepresenting financial information has become a serious concern in the non-profit sector.[1] To qualify for tax-exempt status, non-profits (NPOs) must disclose their finances through tax forms (in the US) or annual returns to the Charity Commission in England and Wales (in the UK). Due to the fierce competition for funding, non-profits have ample opportunities and incentives to mislead potential donors, rating agencies, or the public about their efficiency in accomplishing humanitarian missions. The recent scandals of funding misuses and data manipulation have provoked public demands for increased scrutiny over the sector's transparency. In this chapter, we examine the prevalence of financial misinformation in a representative sample of British non-profits. We revisit a heated debate on the role of information transparency and the trade-off between the cost of accountability and the spending on charitable activities. We investigate whether organisations with higher performance metrics would also report more accurately and whether back-office spending correlates with an increased level of reporting accuracy.

Our paper contributes to at least two active areas of public policy and behaviours of organisations. First, our work is of interest to policymakers and funders who require evaluations of financial transparency of their funding recipients. Previous measures of information misrepresentation based on programme ratios or the distributional discontinuity of financial variables are only applicable to detect irregularities of specific certain financial items (such as investment returns or fundraising costs) from financial reports from a group of organisations (firms) (see Almond and Xia, 2017, Hoffman and McSwain, 2013; Burgstahler and Dichev, 1997). Measures for individual organisations in the for-profit studies are also of limited used due to the lack of systematic organisational information in the non-profit literature (Dechow et al., 2010). Following a revised interest in Benford's Law (Amiram et al., 2015; Barabesi et al., 2017; Villas-Boas et al., 2017), we perform a digital analysis to measure irregularities in financial data for charities in the UK. The underlying intuition is that: to the extent that figures in a non-profit's financial

[1] Chen (2016) documents scandals of fund misuses, abuse of power and lack of transparency in the non-profit sectors. The Centre for Policy Studies, a British think tank, reproduces the accounts of 50 largest UK charities by income and suggests that their actual figures could be more than double the self-reported accounts (Norton, 2014). Keating et al. (2008) find 74% of the regulatory filings from American non-profits fail to properly report categorical expenses, though the authors consider the inaccuracy as due to unintentional errors.

statement are randomly drawn from different sources across several orders of magnitude and not manipulated, the observed distribution of the first digits of these figures is expected to follow a theoretical distribution known as the Benford distribution.[2] The further the observed distribution deviates from the Benford distribution, the more likely the financial figures in the financial statement are interfered by non-random errors.

The method is an appealing alternative for several reasons. First, existing financial data in the public domain is sufficient to run the study. Second, by only focusing on the statistical distribution of the first digits of the financial figures, our approach separates confounding factors such as the scaling effect, forward-looking information or managers' characteristics. Third, the method allows for measuring the deviations from the unobserved true values for each organisation, enabling us to conduct a between-organisation analysis. It is also easily automated and applied to a wide range of economic datasets without specific data selection. Like other measures in the "forensic economics" (Zitzewitz, 2012), the approach is not fail-proof, nor will it substitute existing methods based on auditing. Nevertheless, we believe Benford's Law can serve as a useful and low-cost first step for effective and more targeted investigations and auditing.

Using a large dataset of over 16,000 surveyed public charities and NPOs in the UK, we find that nearly 25 per cent of the sample may have provided misreported financial accounts during the period 2007-2015. To cross-check our results, we referred to the recent charges by the Charity Commission, the governing body of charities in England and Wales. Our analysis correctly identifies 90% of the British charities recently charged for misusing their funding and tinkering with data in 2016. Unlike previous studies using the discontinuity properties of financial figures, we can approximate the degree of potential misreporting for each NPO by the extent to which the observed distribution deviates from the Benford distribution. The proximation allows us to conduct the first systematic study using Benford's Law in the literature of financial disclosure and its determinants. As such, we extend descriptive applications on corporation data (Amiram et al., 2015), macroeconomic statistics (Michalski and Stolz, 2013), and Belgian non-profit data (Van Caneghem, 2015) to understand the reasons underlying NPO misbehaviours.

Our paper also speaks to the dialogue between the public and the charity sector regarding the quest for increased accountability and spending on back-office activities. Keeping accurate financial records is costly and requires financial and human resources

[2] One order of magnitude means about ten times different in quantity than the other. A related distribution but better known in economics is Zipf's law, explained in Gabaix (1999). Both laws are the special cases of Planck's (1901) distribution (see Kafri and Kafri, 2013).

that could otherwise be spent on charitable activities.[3] The opening quote by Singh (2015) asserts the view from the sector: while transparency is a desirable aim, spending on charitable services needs to remain the priority.[4] Without evidence for effectiveness, increased spending on governance tasks could become excessively costly and may distort the organisations' charitable aims (Hofmann and McSwain, 2013; Aldashev et al., 2015).

We show that popular measures such as programme ratios or overhead expenses are not reliable indicators of financial transparency. NPOs with simply either high charitable spending or high governance spending (such as accounting and auditing services) need not report financial information more accurately. We provide a simple model to parameterise the level of misreporting an NPO would commit if it can (mis)report the state of a funded project to a donor before the final impact is realised. Under the assumption that the donor has a prior belief over the misreporting behaviour, the optimal misreporting would depend on both the charitable effort (the trade-off) and the oversight mechanism (spending on governance activities). We test the predictions by examining how the level of misreporting, measured by the Benford digital analysis, is affected by (i) the proportion of income spent on charitable activities, a proxy for the charitable effort, and (ii) the proportion of income spent on governance activities, a proxy for the oversight mechanism. Results from linear regressions controlling for various characteristics and an instrumental variable strategy suggest an interaction effect of the charitable and governance spending on the accuracy of British non-profits' financial data. NPOs with higher spending on charitable activities report more accurately only when spending on governance activities exceeds a certain level (15% of total income in our preferred specification). NPOs with more substantial spending on accounting and auditing services publish financial records more accurately only if their charitable spending exceeds 70% of total income. Inaccuracy appears to be higher when the costs of maintaining accurate financial numbers are higher (for example, in larger and older charities), and lower when the probability of being detected as data manipulator is higher (such as the financial reports being audited or receiving government grants). We complement the IV strategy with a novel estimator by Lewbel (2012) that uses conditional second moments to construct in-data instruments that do not require the standard exclusion restriction. The estimation, together with a battery of robustness checks, provides similar results to the interaction effect.

---

[3] A survey by Woodwell and Bartczak (2008) reveals that 80% of granting institutions did not include sufficient overhead allocations to cover the time and expenses their recipients incurred on reporting requirements.

[4] Singh is the director of the Association of Chief Executives of Voluntary Organisations - a British network for charity and social enterprise leaders.

Our findings bridge several branches of the literature. First, it relates to "forensic economics" studies (Zitzewitz, 2012) on the prevalence and determinants of misreporting. Our application of Benford's Law closely relates to studies using distributional properties of numbers to detect irregularities (Jacob and Levitt, 2003; Michalski and Stolz, 2013; Fang and Gong, 2017). In the non-profit literature, ours is the first systematic paper to consider the manipulation of publicly available financial data of nonprofits (see Hofmann and McSwain, 2013 for a review). By double-checking with charities charged by official bodies, we advocate the use of Benford's Law as a useful investigative tool in the non-profit literature, particularly when the data remain scarce. We also depart from previous non-profit studies by investigating charities' financial records as a whole instead of specific categories. Notable examples include Almond and Xia (2017) showing that some US non-profits manipulate investment returns around zero to avoid revealing negative outcomes. Other studies look at potential misreporting in programme ratios (Trussel, 2003), the levels of social benefits (Vansant, 2016), cost-shifting (Krishnan and Yetman, 2011); or fundraising and administrative expenses (Yetman and Yetman, 2012). Second, we complement studies in the for-profit literature on the determinants of information manipulation (see Bayer et al., 2010 for a review). The majority of the literature suggests that firms could manipulate financial reports to depict a positive financial position in the investment market. Our paper points to two other reasons for misreporting: namely, the trade-off between charitable spending and accounting services and the lack of a strict oversight mechanism. Finally, our theoretical framework provides a different take to explain the misbehaviour of firms and charities. Our model relates to Goldman and Slezak's (2006), Burns and Kedia's (2006), Beyer et al.'s (2014), Thakor's (2015) on designing optimal contracts under potential strategic information disclosure of firms, and Aldeshev et al. (2018) on the misbehaviours of international NGOs. Relatedly, we contribute to the growing literature on the theory of NGO regulation and monitoring (Auriol and Brilon, 2018; Aldashev and Navarra, 2018). Different from these models, we focus on the trade-off and the oversight mechanism as the key explanations for charities' misreporting behaviours.

The paper proceeds as follows. Section 2.2 explains Benford's Law, the UK Third Sector data and how we construct our misreporting proxies based on Benford's Law. We present the theoretical model in Section 2.3. Section 2.4 describes our empirical analysis. Section 2.5 presents the main findings. We summarise various robustness checks to model specifications in Section 2.6 with details in Appendix 2. Section 2.7 concludes.

2.2. Benford's Law and the UK Third Sector Research Data

2.2.1. Forensic Economics Studies and Benford's Law

There are several attempts to measure misreporting in financial data. Popular methods

include estimating accounting models to compare the reported and predicted activities (Dechow et al., 2010) or using discontinuity at zero of earnings or rates of returns to suggest evidence of manipulation (Burgstahler and Dichev, 1997; Lee and Lemieux, 2010).[5] These measures have inherent drawbacks such as correlations with the underlying organisation and manager characteristics (see Amiram et al., 2015), or the confounding effect of scaling, sample selection and research designs (see Durtschi et al., 2005; Gilliam et al., 2015 for the discontinuity approach). These methods are only feasible when detailed records (like the US's IRS forms for non-profits used in Almond and Xia, 2017) or looking-forward information (as in studies using corporate data) are available. Due to this data unavailability, non-profit studies are often able to only look at several categories of financial records instead of the data as a whole.

We improve on the previous literature with a digital analysis based on Benford's Law for each organisation. Benford's Law, also called the first-digit law, is a mathematical law regarding the frequency distribution of the leading digits in many sets of numerical data (e.g., the leading digit of the number 1201.17 is 1). Contrary to basic intuition, the occurrence of each digit as a leading digit in a set of numbers is usually not equal. Instead, the first digits of all numbers in a naturally occurred dataset are expected to occur with a logarithmically decreasing frequency. Physicist Benford in 1938 discovers this pattern and published a series of datasets that adhere to the decreasing distribution. Economist Hal Varian in his 1972 letter to the American Statistics Association promotes the use of Benford's Law in detecting elicit behaviours in economic and financial data. The idea is formalised in Hill's (1999) theorem: For samples randomly taken from a set of numbers following random distributions, the distribution of the first digits of all numbers from these samples will converge toward a distribution called the Benford distribution. The mathematical intuition behind why accurate empirical data would follow Benford's Law is based on three facts. First, the first digit of any number N can be determined by taking its base 10 logarithm and obtaining the fraction behind the integer. For example, if the fraction after the integer of a number N is between 0 and 0.301 (an interval of 0.301), the first digit of N is 1. If the fraction after taking the natural logarithm is between 0.301 and 0.477, the first digit is 2 et cetera. Formally, the intervals between the fractions of the decimal point of the log number ($\log_{10}(1+\frac{1}{d})$) are equivalent to the probabilities that digits appear as the leading number (or digit 1 has a 30.1% of chance). Second, if the probability

[5] For example, Bhattacharya and Tinkelman (2009) examine GuideStar data of 111,000 non-profits by distributional analysis and find no evidence of expense allocation manipulation. Similarly, Ballantine et al. (2007) find a highly significant discontinuity in residual incomes of English NHS hospitals during 1998-2004.

distribution function of the logarithm of N is smooth and symmetric, a number will be in the internal between n and n + 0.301, where n is an integer in the logarithmic distribution, with a probability of 30.1%, between n + 0.301 and n + 0.477 with 17.6% et cetera. The implication is that for numbers without human manipulation, there is a 30.1% (17.6%) of chance that their first digits would be 1 (2, respectively). Third, according to the Central Limit Theorem, distributions drawn from a random mixture of different distributions would be smooth and symmetric. The implication of is that sets of data that comprise of different sources of numbers would have a smooth and symmetric probability distribution function such that the first digits of all of the numbers should follow the Benford distribution. Hill's (1999) theorem provides the following formal derivation:

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

where $P(d)$ is the probability that digit $d = 1, 2, \ldots, 9$ occurs as the leading digit. Table 2.1 records the theoretical distribution specified by Benford's Law: 1 will appear as the leading digit 30.1% of the time, 2 will appear 17.6% of the time, and so forth.

Table 2.1. 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

The conditions laid out in Hill's (1995) theorem are likely to apply to accurately reported financial data for two reasons.[6] First, the correct (unobservable) realisations of financial items in the financial reports, such as total revenues, revenues from different sources or cash flows, are determined by many interactions by many individuals during a given period. These interactions could be considered as randomly distributed since they are known only to those involved. The financial items representing these interactions, therefore, are likely to be governed by different mechanisms. For example, the distribution of revenues from government funding plausibly differs from that of administrative costs. The mixture of these estimates, which constitute an organisation's financial report, would follow Hill's Theorem. Specifically, the aggregated set of numeric items representing revenue sources from grants, businesses or investments, together with expenditure figures on salaries, charitable activities, taxation, is expected to follow Benford's Law.

Following Nigrini's (1996) seminal paper, the idea of using Benford's Law to detect manipulations in financial data is now frequently mentioned in auditing and

[6] See Villas-Boas et al., 2017 and citation within for recent statistical evidence of the law's applicability to economic behavioural micro-data.

accounting.[7] Nigrini (2012) documents successful application of the method, including his detection of fraudulent financial reports of seven companies in New York City (commissioned by authorities of the City of New York). Durtschi et al. (2004), (Nye and Moul, 2007), Michalski and Stoltz (2013), Miller (2015), and Amiram et al. (2015) discuss how auditors can effectively use Benford's Law in detecting errors and frauds in annual reports and macroeconomic data. The main consensus is that accounting-related data are expected to adhere to the Benford distribution and as the deviation from the Benford distribution increases, the degree of errors increases. Amiram et al. (2015) provide the first simulation analysis using stylised financial statements to ascertain this property of Benford's law. They show that only after introducing non-zero mean errors to the dataset do they see deviations from the Benford distribution; and the larger the error introduced, the larger is the deviation from the law. This property is akin to the idea of "hard-to-forge" signatures (Kossovosky, 2015 p.109). We exploit this characteristic to construct our measures for each charity below.

2.2.2. Data

We use the Third Sector Research data deposited in the UK Data Services by Alcock and Mohan (2015) as the representative dataset for the British charity sector. The data are collected in five phases (first by the Third Sector Research Centre 07-08; and then by the UK Civil Society's Almanac in 2012, 2013, 2014, 2015) and include yearly financial statements of 16,391 charities for the period 2007-2015 (up to eight annual reports for each NPO).[8, 9] For each phase, financial characteristics were extracted from information on *all* registered charities in each year held by the Charity Commission for England and Wales, the Offices of the Scottish Charity Regulator, and the Financial Conduct Authority. Apart from standard items in a financial statement, the dataset provides detailed financial information on numerous types of expenses such as charitable and fundraising activities, voluntary incomes, administrative expenses, and employment statistics (see Appendix

[7] Following Zitzewitz's (2012) taxonomy, measures based on Benford's Law can be categorised as a statistical model-based approach. The primary assumption is similar to ours: fraudulent cases exhibit patterns that are very unlikely under a statistical model of honest behaviours.

[8] Due to survey design, only non-profits with the total income of at least £25000 are collected.

[9] We first convert all reported financial items to Sterling using relevant exchange rates. There are charities whose headquarters are in Britain but operate abroad and choose to report in the local currencies (euros, Thai baht, Singaporean dollars, US dollars). The conversion does not alter the conformity of the dataset due to the scaling invariance property of the Benford distribution (see Morrow, 2014 or Michalski and Stoltz, 2013 for proof).

Table 2.A1 for a list of financial items provided). Absent of errors, such variety of financial items available increases the chance of conformity to the Benford's Law as these are drawn from different types of underlying distributions such as revenues, donations, or expenses.[10] All the figures are recorded rather than constructed from raw data (as an index). This feature preserves the "naturalness" of the underlying distributions as it avoids seasonal adjustments or statistical methodologies used in constructing an index.

For objectivity, we use all the financial information that appears on the statements to calculate proxies of reporting inaccuracy for each charity in our sample. In a standard financial statement requested during the surveys, there are 135 entries. However, many are recorded as zeros. Since in the raw dataset, there is no coding for missing observations we treat these zeros as genuine information, that is, transactions whose values are zero. This type of recording is a challenge for our analysis as the rule-of-thumb requirement for constructing this type of measures is that it requires at least 100 non-zero quantities in each tested unit.[11] Ideally, we would construct organisation-year indices, as in Amiran et al. (2015), to study the dynamic natures of information irregularities. Due to data availability, however, we follow Michalski and Stolz (2013) and Henselmann et al. (2012) to pool each NPO's annual reports over the available years. This approach particularly suits our analysis since NPOs in the surveys provide their annual financial statements by waves. That is, most of the NPOs provide several consecutive financial statements in one round. This practice allows the NPOs to inject errors in multiple annual accounts and it is appropriate to consider these multiple annual accounts at once. To alleviate concerns with yearly pooled data, we report in Section VI a robustness check when we split the sample into two periods (pooling four years of data together). Since the robustness check of pooling four years of data would severely reduce our sample, we consider a cross-sectional sample of 10,322 charities that provide at least 100 non-zero financial figures in our main analysis.[12]

---

[10] We remove NPOs with negative assets or spending on governance or charitable activities.

[11] As the number of digits N goes to infinity, the distribution converges to a chi-square distribution with 8 degrees of freedom. The chi-square statistics converges to the limit when $N \geq 30$ and $NP_e(d_i) \geq 5$ for all $d_i \in \{1, \ldots, 9\}$. That is, the lower bound for N to be a valid statistic is $N \geq \frac{5}{P_e(d_i=9)} > 100$. Milchaski and Stoltz (2013) run simulations and suggest that tests for Benford's Law is powerful only for sample with at least 110-digit points. We experiment with both the cut-offs and find similar results.

[12] Pooling yearly data is not uncommon in digital analysis based on Benford's Law. Amiran et al. (2015) through simulated analysis and comparing with existing measures of reporting quality show that non-fabricated annual financial statements, whether in aggregate, by year, or by

We remove NPOs whose total number of non-zero financial items after pooling is fewer than 100. Conceptually, excluding this part of the sample could lead to a selectivity bias in that some non-profits may choose to enter zero entries in each report strategically. We show that this bias does not drive our results. First, we replicate our main results when varying the threshold of at least 100 non-zero figures from 65 to 115. Second, zero transactions in an annual report could reflect an NPO's choice either not to participate in some activities or to withhold information. Non-participation is not a serious concern: it is plausibly independent of manipulative behaviour because the two mechanisms governing the decisions are different. We cannot address the information withholding concern with this dataset. We argue, however, that it is not critical to our analysis for three reasons. First, as the balance sheet in each financial year must remain in balance, withholding information by recording some transactions as zeros would require manipulating other non-zero financial items.[13] Our measures based on Benford's Law would pick up these misreporting from the non-zero items. Second, we include in our empirical analysis a variable specifying the number of non-zero financial observations used in constructing the measures. The variable aims to account for both the diversity of the NPO's activities and, potentially, the level of the NPO's intention to disclose their financial details. Third, we estimate a sample selection model for missing observations using a Heckman correction model to distinguish between two decisions: (1) whether to report more than 100 non-zero financial items (or to be included in our analysis); and (2) the optimal degree of inaccuracy conditional on at least 100 non-zero items are reported. Despite being tentative due to the lack of detailed data, we obtain similar results.

---

organisation-year are expected to conform to Benford's Law. Nigirini (2011, chapter 17) also uses multi-year financial statements to demonstrate the applicability of aggregate data in assessing errors and frauds by digital analysis. Henselmann et al. (2012) showed that Benford's law applies to the aggregated data of single-company annual reports given their reports are free of errors. Nye and Moul (2007) show theoretically that pooled macroeconomic statistics conform to Benford's Law. Michalski and Stoltz (2013), also citing the lack of detailed data, aggregate quarterly macroeconomic data from several countries according to their economic characteristics. Using random subsampling to draw Bernoulli random subsamples from the aggregated data subsets, they show that their whole dataset adheres to Benford's Law. Kossovsky (2015, p.90) provides a detailed review of the previous literature to suggest a collection of monthly/yearly or the market aggregate (combining multiple companies) data is also expected to conform to the law.

[13] There is no recorded information for balancing errors in the raw dataset.

2.2.3. Measures of the level of reporting inaccuracy

There are two popular methods to measure inaccuracy using Benford's Law: (1) using a measure of statistical dispersion (such as the Median Absolute Deviation), and (2) using test statistics and critical values to establish (reject) the conformity of the tested distribution (see Amiram et al., 2015). The main concern of using the second method is that test statistics are sensitive to the number of digits used. When the number of digits used increases, test statistics tend to over-reject the null hypothesis of the observed distribution adhering to the Benford distribution. It is because the critical values for these tests increase with the sample size (the number of digits used) that they require perfect conformity to establish (fail to reject) the null. The first method avoids this concern. The measure of statistical dispersion does not require a critical value, providing an objective comparison across organisations with different numbers of digits tested (Nigiri, 2012). For these reasons, we use the Mean Absolute Deviation (MAD) statistic in the main analysis.[14] The MAD is calculated as the mean of the absolute difference between the empirical proportions of each digit in each NPO's pooled financial reports and their respective theoretical proportion according to Benford's Law (see Table 2.1):

$$\text{MAD} \equiv \frac{1}{9}\sum_{i=1}^{9}|P_o(d_i) - P_e(d_i)| \tag{2}$$

where $d_i = 1, 2, \ldots, 9$ represents the digit; $P_o(d_i)$ is the observed proportion of digit d_i, $P_e(d_i)$ is the expected proportion of digit d_i according to Table 2.1. To interpret, the larger the MAD statistic, further the deviation from the theoretical distribution under the null hypothesis that the pooled report is free of errors and misrepresentation.[15]

In Section 2.6, we rerun the analysis with the three "critical value based" measures created from (1) the Pearson's chi-square test statistics (χ^2) of goodness of fit, (2) the Kolmogorov – Smirnov (KS) statistics and (3) a binary variable (*Deviate)* of whether we reject the null hypothesis of the data conforming to the Benford distribution using KS tests at the significance of 5% (1 = Yes, 0 = No). Despite the drawbacks discussed above, these measures have been widely used by previous studies and practitioners due to their ease of

[14] See Morrow (2014) and Miller (2015) for detailed discussion. Measures can be strongly influenced by the number of digits used, with some statistics requiring near-perfect conformity to the theoretical distribution as the number increases to not reject the null of conformity (Nigrini, 2012). Barabesi et al. (2017) propose a testing procedure that claims to deliver exact significance levels and not rely on large-sample approximation.

[15] Nigrini (2012) recommends a table of "critical values for rejecting conformity" for practitioners. However, it is based on simulated datasets of specific dataset types.

use and practical interpretations. Like the MAD statistic, higher values of the test statistics show the tested data diverge farther from the Benford distribution. Examples of the Chi-square test include Nye and Moul (2009), Michalski and Stoltz (2013); and those of the KS test include Morrow (2014), Amiran et al. (2015). The Pearson Chi-square statistic is the simplest measure to investigate whether distributions of two categorical datasets differ from each other. The KS statistic quantifies the cumulative distance between the observed distribution of the tested organisation and the reference distribution (here, the Benford distribution). For our analysis, we calculate the KS statistic as the maximum deviation from the Benford distribution. Finally, we use the binary *Deviate* for descriptive purposes to examine the number of organisations that provide an accurate set of financial data. The specific constructions of the test statistics are as below:

$$\chi^2 \equiv N \sum_{i=1}^{9} \frac{[P_e(d_i) - P_o(d_i)]^2}{P_e(d_i)} \tag{3}$$

$$KS \equiv \max_{d_i \in \{1,2,\ldots,9\}} |\sum_{i=1}^{d_i} (P_o(d_i) - P_e(d_i))| \tag{4}$$

$$\text{Deviate} = \begin{cases} 1 \quad \text{if} \quad KS \leq D_N(\alpha) = \dfrac{c(\alpha)}{\sqrt{2N}} \\ 0 \quad \text{if } KS > D_N(\alpha) = \dfrac{c(\alpha)}{\sqrt{2N}} \end{cases} \tag{5}$$

where N is the total number of non-zero financial items used, $D_N(\alpha)$ is the critical value of the Kolomogorov distribution at N and test power α, $c(\alpha) = \sqrt{-\frac{1}{2}\ln(\frac{\alpha}{2})}$ is the Benford specific critical value at α calculated in Morrow (2014). Normally. for a significance level at $\alpha = 0.05$, $c(0.05) = 1.48$.

Caveats of our measures: These measures hinge on the premise that an accurate financial statement dataset adheres to Benford's Law; while manipulated and erroneous data deviate from the law. Several factors may complicate our analysis. First, an NPO may not cheat all the time (serial cheaters might be found out more quickly) thus the measures based on pooled data cannot pinpoint for which year or which financial items the illicit behaviour have occurred. Second, the measures cannot detect other subtle types of cheating: such as rounding up numbers or petty manipulations which also only affects the last digits (for example recording £1500 instead of £1268). These types of manipulation are difficult to deal with and require much richer data (see Schennach, 2013 for recent research). Third, one organisation could change all financial items by a common factor or in a creative way that preserves the Benford first digit distribution. Since changing one first digit of entry would later require altering other entries' first digits, we expect that this manipulation is costly to implement. The experimental literature also shows that people tend to badly replicate known data-generating processes even when instructed to do so (Camerer, 2003,

pp. 134 – 138). As Benford's Law is widely used in the professional services but not publicly well-known (Cho and Gaines, 2007), it is unlikely that organisations would be able to preserve the Benford distribution. Bearing that in mind, we theoretically address the cost of, if any, preserving the distribution by introducing the governance cost that captures the effort of the agent to manipulate the report creatively. Empirically, we argue that it is not critical because measurement errors in the dependent variable (here the misreporting measure) do not lead to biased estimates. The only consequence is less precision in the estimated coefficients and lower t-statistics (Hausman, 2001). Fourth, deviations from the Benford distribution could be due to poor data collection/bookkeeping (human errors) without an intention to mislead regulatory bodies. Although we cannot rule out the possibility of errors, we doubt that human errors could drive the deviations. First, rounding the first digit is rare (except for cases such as rounding £1998 to £2000). Second, if the rejection of the Benford distribution were caused by poor bookkeeping and data maintenance, these would be NPOs with the lowest spending on governance activities (such as accounting or auditing). The data do not support this implication: considering NPOs with the lowest 10% of spending on governance activities, for 35% of these NPOs we fail to reject that their financial reports do not adhere to the Benford distribution according to the KS tests at the 5% level of significance. Our results remain consistent when we exclude NPOs in the bottom 10%, 25%, and 50% of spending on governance. We conjecture that if poor bookkeeping resulted in fewer data points being collected, including these NPOs would alter our results. Our results hold when we alternate the threshold of the number of non-zero financial items used (see Appendix 2.8.5). Acknowledging these caveats, we model the interaction between charities and the governing body to explain irregularities in public reports in the next section.

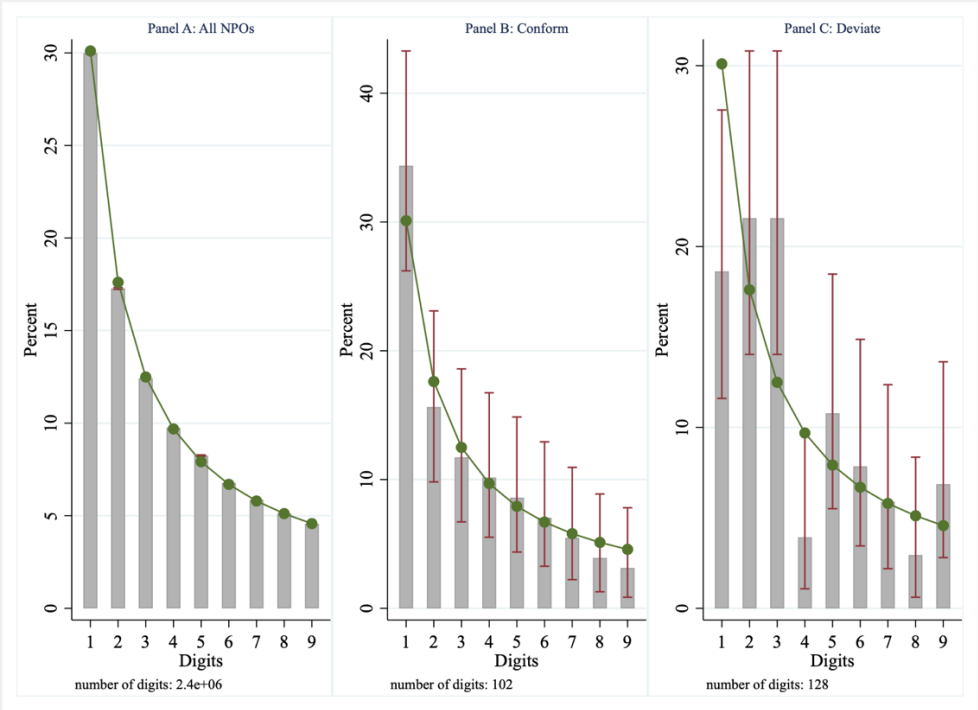
2.2.4. An illustration of Benford's Law

Figure 2.1 shows graphical evidence to support the applicability of Benford's Law to our data. When combining available financial figures from all the surveyed charities, the distribution of the first digits of these figures closely follows the Benford distribution, assuring our premise that accounting data should follow the law.[16] When each charity is

[16] The conformity of the data does not prevent the possibility that some individual charities may have inaccurate financial data. It is because the overall conformity may come from a mixture of independent errors embedded in different charities' data (different manipulators might manipulate different items in different ways). 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.

considered, we fail to reject the null hypothesis for the remaining 75% of the sample with complete disclosure. Panel B provides a representative distribution of this group. We consider these NGOs as "conforming" to the law, showing that we fail to find evidence, both statistically and graphically, of potential misreporting. In contrast, we reject the hypothesis that the observed distribution of the first digits of all the numbers follows the theoretical distribution for 25% of the sample with complete disclosure. We call these charities as "deviating" from the law, suggesting that their full financial accounts may contain inaccuracy detectable by the law. Panel C provides an example of a charity charged by the Charity Commission of England and Wales in late 2016 for tinkering with data. For anonymity agreement, we remove the charity's name. Panel C shows clear a graphical deviation from the Benford distribution, suggesting potential data manipulation.

Figure 2.1. The UK Third Sector Research Data and the Benford distribution



Note. Lines represent the theoretical distribution in Table 2.1. Bars 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 all the NPOs. Panel B is for a representative NPO which we fail to reject the hypothesis test of conformity to the law using the Kolmogorov – Smirnov test (75% of the sample). Panel C is for a representative NPO whose requested financial accounts fail the hypothesis for its requested financial accounts (25% of the sample). For a representative purpose, we use one of the charities charged by the Charity Commission of England and Wales in late 2016.

2.2.3. A theoretical model of optimal misreporting

Consider a three-period reporting game between a donor (principal) and an NPO (agent) over a funded project. Our setup is like Goldman and Slezak's (2006) where the agent may take a hidden action, which affects the (actual) terminal value of the project, and potentially misreport the intermediate value (such as financial records of the organisation) to the uninformed donor to gain a higher payoff. In our context, the payoff can be either periodic grant disbursements subject to satisfactory performance reports of the agent or future grants that use the report as part of the fundraising application. As such, the agent has an incentive to inflate the unrealised value in their report to the donor.[17] The optimal contract is to incentivise the agent to work on the project's actual value and to minimise the agent's incentive to misreport the value. Different from their model, we distinguish two types of reporting errors: intentional manipulation and unintentional errors such as failing to comprehend and estimate the current state of the organisation or simply human errors when recording information. For simplicity, we assume the human errors (genuine mistakes) in the bookkeeping process are specific to organisations and unknown to the agent when reporting their value. We discuss key features of our model below.

2.3.1. Basic building blocks

At $t = 0$, a risk-neutral donor contracts with an NPO to deliver a social project that yields a terminal value in the long run $t = 2$. The NGO is assumed risk-averse since they are not allowed to distribute their profit (see Wedig, 1994). During $t = 0$, the NPO privately makes two one-time decisions. First is the amount of unobservable action $a \geq 0$ (such as the level of dedication or effort). The second is the extent of misreporting (shortly as errors, denoted b) of the report that the NPO will issue at $t = 1$ (such as how much the report will inflate the privately observed intermediate state of the project). Exerting effort and misreporting are both costly to the NPO. Let the NPO's disutility of exerting action a be $\psi^a(a) = \frac{\delta}{2}a^2$, where the convex functionality represents the increasing marginal disutility at rate $\delta > 0$. Let $\psi^b(b) = \frac{g}{2}b^2$ represent the NPO's cost of producing a report with an amount (b) of misreporting, where the *governance spending* $g > 0$ is the spending on governance/auditing activities by the agent. The cost of misreporting $\psi^b(b)$ reflects the disutility of misreporting and has two components. First, a higher level of misreporting b leads to an increasingly higher level of disutility due to two factors: (i) the time spent lobbying the auditor or

[17] This assumption is consistent with manipulation incidents documented in Krishnan and Yetman (2011) that Californian non-profit hospitals may misreport program ratios to the state regulatory agency by +8%.

coming up with creative ways to go around monitoring requirements and (ii) the intrinsic aversion to lying (Hurkens & Kartik, 2009; Gneezy, 2005). Second, a higher level of governance cost g represents a higher governance quality in the organisation, leading to a higher disutility caused by intentional manipulation of the report. By governance quality, we follow Beyer et al.'s (2014) and Thakor's (2015) interpretation as the oversight mechanism required by the donor, for pressurising the agent to conform to accountability and aligning the NPO's interests more closely with the donor's (for example the composition of the oversight board/committee; or the NPO's accounting division and choice of external auditors). In practical terms, governance spending involves administrative expenses, auditing and accounting fees that are directly observable in our data.

Finally, we assume the NPO incurs a reputation loss $\psi^c(b) = c(b - b^e)$ for deviating from the donor's prior belief of the NPO's equilibrium level of misreporting $b^e \geq 0$. The linear functionality is for tractability and captures two notions. First, the amount of the reputation loss $\psi^c(b)$ is increasing in the actual amount of misreporting, regardless of the prior belief $b^e \geq 0$. Since the NPO only suffers any reputation concern if the principal finds out the agent deviates from the principal's prior belief of the misreporting intensity, we use parameter c to capture the probability that the principal detects the deviation. The higher the probability of being found out, the higher reputation loss (gain) is when the NPO deviates from (conforms with) the donor's prior belief. Taking this probability into consideration, if the NPO's misreporting choice is *better* than expected, that is the NPO misreports less than what the donor expects $b < b^e$, the NPO's reputation gains $\psi^c(b) > 0$. Otherwise, the NPO's reputation reduces by $\psi^c(b) < 0$. Second, the linearity and the parameter $c \geq 0$ reflect how deviating from some prior belief of reputation matters to the NPO for future fundraising activities. Empirically, we can think of c as capturing the degree of repetitive interactions between the donor and the agent, for example, whether the report is subject to external auditing or the degree that the NPO's income relies on resources from fundraising or grant applications. Intuitively, NPOs without large internal funds (such as endowments or inherited grants) would have to rely on external supports and would have the incentive to maintain a strong impression with the donor, thereby minimising the level of misreporting. We will test how these factors impact misreporting below.

2.3.2. Timeline

At $t = 0$, the agent chooses action a that yields a gross terminal value at $t = 2$, denoted as $V = \rho a + \eta + \varepsilon$. Parameter $\rho > 0$ is a productivity factor, $\eta \sim N(0, \sigma_\eta^2)$ reflects random organisation-specific uncertainty, and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ represents random idiosyncratic shocks faced by the NPO after the initial period. As discussed above, the organisation-specific uncertainty captures η two features. First, there are organisation shocks to the NPO such

as termination of funding towards its main services. Second, it captures human errors generated during the bookkeeping process. These errors are assumed random and unknown to the NPO when deciding the optimal action a.[18] Regarding the information set, we assume η and ε are unknown to the NPO when choosing the level of action a at $t = 0$; while ρ, σ_η^2 and σ_ε^2 and the parameters of the cost functions are commonly known.

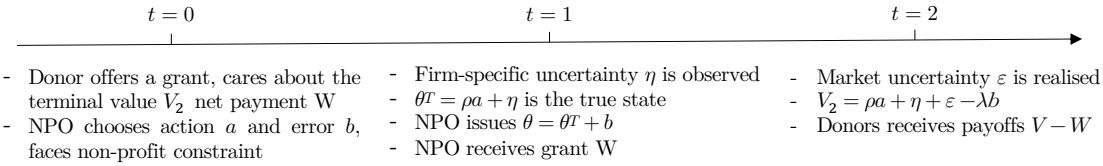
Since the true intermediate value is unobserved by the donor, NPO could misreport by an amount of errors b and issues the report $\theta = \theta^T + b$. Assume for convenience that $b \geq 0$ or the NPO tends to over-report the project's value to mislead the donor for a higher payoff. Based on the observed report, θ, the donor forms their expected terminal value, S, and disburses the contracted grant W specified below.

At $t = 2$, the actual gross terminal value V induced by action a and the amount of inaccuracy b are recognised. Recall that at $t = 0$, by choosing the misreporting amount b, the NPO diverts some of the project's resources away from productive uses to prepare for the misstated report at $t = 1$ (e.g., monetary cost for bribing or colluding with auditors, or the opportunity cost of the NPO's time spent on manipulating the accounts). For simplicity, we assume this diverted resource linearly reduces the project's terminal value. Namely, for an amount of errors b, the gross terminal value at $t = 2$ falls by λb, where the commonly known $\lambda > 0$ parameterises the incremental cost of the resources diverted. The net terminal value induced by action a and manipulation b is given by:

$$V_2 = V - \lambda b = \rho a + \eta + \varepsilon - \lambda b = \theta^T - \lambda b + \varepsilon \tag{6}$$

Figure 2.2 summarises the model's timeline.

Figure 2.2. Timeline of the three stages.



2.3.3. Payoffs

The first-best solution occurs when the donor could contract on the project's terminal value so that there is no misreporting. Given the value often takes time to recognise, it is unrealistic to wait until the terminal value is observed before the NPO is compensated.

Instead, we focus on the second-best solution in which there exist hidden action and information. The donor needs an indication of the project's terminal value to pay the

[18] An extension to this model is to endogenise the human errors. That is, instead of assuming a random process, we assume that human errors can be reduced by increasing the level of spending on accounting and auditing activities.

NPO according to their performance at $t = 1$. The expected value $S = E(V|\theta)$, based on the agent's report θ, is the only observable performance measure the donor can use. As standard, we assume that at $t = 0$ the donor designs a linear contract $W(S)$ as:

$$W(S) = w_o + w_1 S \tag{7}$$

where w_o is the upfront payment and w_1 is the value-sensitivity of the contract.[19]

The donor knows of potential misreporting in the report θ and forms their belief of the net terminal value V_2 by subtracting the received report θ by an amount b^e: $V_2 = \theta - b^e$. Similar to (Stein, 1989) and Golman and Slezak (2006), we assume that this prior belief b^e is formed before the contract begins, remains exogenously fixed due to the one-off nature of the interaction, and is not updated after observing the report θ.[20] This belief could be formed by examining previous records of the NPO and their organisational structure. When this belief is rational, that is the donor perfectly predicts the equilibrium level of misreporting resulting from the optimal contract $b^e = b^*$, information misreporting has no impact on the expectation of the gross terminal value.[21] Otherwise, the expected value S is increasing in the actual amount of misreporting b and decreasing in the expected intensity of the NPO's misreporting b^e. Formally, the expectation of the net terminal value given the report θ at $t = 1$ is:

$$S = E[V_2|\theta] = E[\theta^T + \varepsilon|\theta] = \theta - b^e - \lambda b^e = \rho a + \eta + b - b^e - \lambda b^e \tag{8}$$

At $t = 1$, the NPO undertakes the contract, receives $W(S)$, incurs the disutility of effort, misreporting and reputation loss: $\psi^a(a) = \frac{\delta}{2}a^2$, $\psi^b(b) = \frac{g}{2}b^2$ and $\psi^c(b) = c(b - b^e)$, respectively. The induced wealth of the NPO is:

$$\omega = W(S) - \psi^a(a) - \psi^b(b) - \psi^c(b) \tag{9}$$

As standard, we assume the NPO has a constant absolute risk aversion (CARA) utility function, $u(\omega) = -\exp(-r\omega)$ with the Pratt-Arrow absolute risk aversion coefficient of $r > 0$. According to the zero-profit assumption, we normalise the reservation utility u_0 to zero. With the payoffs specified, there are several standard results.

---

[19] The term "value-sensitivity" means the grant paid to the agent is linearly correlated with the reported mid-term value. For example, if the agent reports they have reached out 1,000 additional beneficiaries, the donor will pay accordingly £10,000 more. Here, $w_1 = 10$.

[20] Another extension to assume that the donor can be naïve and expect that $b^e = \tau b^*$ with probability of τ; while the donor can be sophisticated and perfectly expect that $b^e = b^*$ with probability of $(1 - \tau)$. The empirical predictions of interest remain.

[21] One complex extension is to assume a Bayesian game with updating beliefs and punishment in dynamic interactions. We chose not to model such a game for parsimony. One example is Benabou & Laroque's (1992) study on information manipulation in financial market.

Lemma 1. Given NPO's information set ω_0^N at $t = 0$, the NPO would require a certainty equivalent for undertaking contract W whose induced wealth at $t = 1$ is given by (9) as:

$$CE(W, a, b|\omega_0^N)$$

$$= w_0 + w_1(\rho a + \eta + b - b^e - \lambda b^e) - \frac{\delta}{2}a^2 - \frac{g}{2}b^2 - c(b - b^e) - \frac{r}{2}w_1^2\sigma_\eta^2 \tag{10}$$

Proof. See Appendix 2.8.1.

The term $\frac{r}{2}w_1^2\sigma_\eta^2$ reflects the premium that the NPO needs to protect themselves against organisation-specific shocks η, which is not realised until $t = 1$.

At $t = 0$, given the information set $(w_0, w_1, \rho, b^e, \lambda, \delta, g, r, \sigma_\eta^2)$, the NPO aims to maximise their utility by choosing action a and the misreporting level b. As the expected utility is equivalent to the utility at certainty equivalent $E[u(\omega)|\omega_0^N] = u(CE)$ and the utility function is monotonic, the NPO's problem is equivalent to maximising the certainty equivalent regarding a and b:

$$\max_{a,b} CE = w_0 + w_1(\rho a + \eta + b - b^e - \lambda b^e) - \frac{\delta}{2}a^2 - \frac{g}{2}b^2 - c(b - b^e) - \frac{r}{2}w_1^2\sigma_\eta^2 \tag{11}$$

The first-order condition gives the NPO's action and misreporting choice as:

Corollary 1. The NPO optimally responds to the contract $W(S) = (w_0, w_1)$ by choosing:

$$a^* = \frac{\rho}{\delta}w_1 \tag{12}$$

$$b^* = \max\left\{0, \frac{w_1 - c}{g} + \tau\right\} \tag{13}$$

As standard, to induce the agent to exert any action, the principal must offer a contract positively sensitive to the performance measure S $(w_1 > 0)$; while the optimal action is decreasing with the marginal rate of disutility δ. When the reputation concern dominates the marginal benefit from the performance-based sensitivity, the agent's optimal strategy is to report truthfully. Namely, if the NPO wants to maintain a good record or future contract with the donor, they will report truthfully. If the incremental reputation concern c is sufficiently small, there exists a positive value of optimal inaccuracy. For the rest of the analysis, we examine this situation where $w_1 > c - \tau g$.

The donor considers the optimal levels of action and misreporting to design a contract $W(S)$ that maximises the terminal value net the grants by solving:

$$\max_{w_0, w_1}\{V_2 - W(S)\} \tag{14}$$

subject to the incentive compatibility $\{a^*, b^*\} = \arg\max_{a,b} \text{CE}$ and the participation constraint:

$$CE = w_0 + w_1(\rho a + \eta + b - b^e - \lambda b^e) - \frac{\delta}{2}a^2 - \frac{g}{2}b^2 - c(b - b^e) - \frac{r}{2}w_1^2\sigma_\eta^2 \geq u_0 = 0 \tag{15}$$

In equilibrium, the participation constraint holds at equality so that the NPO earns zero profit. $CE = 0$ implies that:

$$w_0 = -\left[w_1(\rho a^* + \eta + b^* - b^e - \lambda b^e) - \frac{\delta}{2}a^{*2} - \frac{g}{2}b^{*2} - c(b^* - b^e) - \frac{r}{2}w_1^2\sigma_\eta^2\right] \tag{16}$$

Substituting w_0, the optimal action (12) and misreporting (13) into (15), we solve:

$$\max_{w_1} \frac{\rho^2}{\delta}w_1 - \lambda\left[\frac{(w_1 - c)}{g} + \tau\right] - \frac{\rho^2 w_1^2}{2\delta} - \frac{(w_1 - c)^2}{2g} - c\left[\frac{w_1 - c}{g} - b^e\right] - \frac{r}{2}\sigma_\eta^2 w_1^2 \tag{17}$$

The first-order condition gives the unique equilibrium contract (w_0^*, w_1^*) as:

Corollary 2. There exists unique (w_0^*, w_1^*) such that w_1^* satisfies (17) and w_0^* follows (16):

$$w_1^* = \frac{\frac{\rho^2}{\delta} - \frac{\lambda}{g}}{\frac{\rho^2}{\delta} + \frac{1}{g} + r\sigma_\eta^2} \tag{18}$$

i. The value-sensitivity w_1^* decreases with the disutility of effort δ, risk aversion r, and the agent's specific uncertainty at the intermediate state σ_η^2

ii. The value-sensitivity increases with and the agent's productivity ρ

iii. The lump-sum amount w_0^* is set for the reservation utility equals to zero.

We assume that *governance spending* g is sufficiently high, $g > \frac{\delta\lambda}{\rho^2}$, so that the performance-based sensitivity w_1^* is always positive. As such, the equilibrium performance-based sensitivity is a function of the diverted resource cost λ, the governance spending g, the incremental compensation for productivity ρ, action cost δ and the organisation-specific risks borne by the NPO $r\sigma_\eta^2$. Equation (18) specifies the form for the contracted value sensitivity. Implications 2 are standard: the donor will always set a positive performance-based sensitivity w_1^* to incentivise productive effort. The sensitivity will be higher for a higher-productivity agent while being lower for an agent with less incentive to work (higher disutility of effort) or with a higher level of risk aversion or associated risk. Implication (iii) is equivalent to the zero-profit assumption of the NPO.

Substituting w_1^* into (13), the optimal amount of misreporting now becomes:

$$b^* = \frac{\beta - \frac{\lambda}{g}}{g\beta + g\Delta + 1} - \frac{c}{g} + \tau \tag{19}$$

where we define $\beta = \frac{\rho^2}{\delta}$ and $\Delta = r\sigma_\eta^2$ for convenience.

2.3.4. Comparative statics and testable predictions

The following propositions specify comparative statics that characterises the equilibrium interactions between the optimal amount of action and misreporting regarding observable characteristics. We restrict the results to those needed for the empirical analysis.

Proposition 1: When the exogenously imposed governance pressure is sufficiently high for the donor to offer a positive performance-based sensitivity $w_1^* > 0$, there exists a fixed threshold of the *governance spending* $\tilde{g} > 0$ such that:

$$\tilde{g} = \arg\max_{g} b^* = \frac{\lambda + c + \sqrt{(\lambda + c)\lambda + \frac{\beta(\lambda + c)}{\beta + \Delta}}}{\beta - c(\beta + \Delta)} \tag{20}$$

where $\beta = \frac{\rho^2}{\delta}$ and $\Delta = r\sigma_\eta^2$ and:

i. $\frac{\partial b^*}{\partial a^*}|_{\partial g} > 0$ if and only if $g < \tilde{g}$

ii. $\frac{\partial b^*}{\partial a^*}|_{\partial g} < 0$ if and only if $g > \tilde{g}$

Proof: See Appendix 2.8.2. Intuitively, the threshold effect follows the non-monotonicity of the optimal misreporting function because the governance spending term enters the optimal misreporting function non-monotonically.

Proposition 1 suggests a high level of charitable effort needs not be a signal for the report being more accurate. Part (i) suggests that if an NPO's governance spending falls below a certain threshold, agent exerting a higher level of productive effort will report the intermediate state less accurately. If the governance spending exceeds the threshold, a highly productive agent will be more likely to issue an accurate report. There are two intuitive explanations for this proposition. First, a low level of governance spending will impose a looser accountability mechanism over the agent's reporting procedure, allowing the highly productive agent to inflate their reported state to capture an even higher level of payoffs. In contrast, when the governance spending is sufficiently high, the stricter governance mechanism prevents manipulative behaviours of the highly productive agent. The reason is that misreporting becomes too costly for the highly productive agent: under strict scrutiny, "good" organisations would issue accurate statements to avoid potential punishments associated with being detected as untruthful (Benabou and Tirole, 2006). Second, a sufficient level of spending on administrative and accounting activities may reduce human errors, particularly when the agent focuses on generating highly productive effort. Whereas when the level of productive effort is low, the agent will divert the resources to creative accounting and intentional misreporting, increasing the intensity of misreporting which is already prone to human errors.

Proposition 2: When the governance spending is sufficiently high so $w_1^* > 0$, there exists a fixed threshold of the optimal action $\tilde{a} > 0$ such that:

i. $\frac{\partial b^*}{\partial g} > 0$ if and only if $a^* < \tilde{a}$

ii. $\frac{\partial b^*}{\partial g} < 0$ if and only if $a^* > \tilde{a}$

Proof: See Appendix 2.8.3.

Proposition 2 suggests that higher spending on governance activities (a tighter oversight mechanism) needs not guarantee reports to be more credible. Agent with higher governance spending would still report the intermediate state less accurately if their optimal effort falls below a threshold $\tilde{a}$ (a low type). In contrast, if the optimal effort is

exceptionally high, agent with a tighter oversight mechanism would report the state more closely to the true value. One way to intuitively explain Proposition 2 is to classify two types of agents: high (low) type NPOs are those who choose their optimal effort higher (lower) than the threshold (for example, an industry norm, or implicit agreement with donors). Low-type NPOs could divert the spare effort and resources to devising creative accounting to inflate the intermediate state. Meanwhile, high-type NPOs are now constrained by the limited resources, which have been spent on the project, and would choose to report accurately for two reasons. For one, reporting accurately is now cheaper than devising creative techniques to overcome the tighter oversight mechanism. Second, high-type NPOs are more concerned with their reputation, especially when the tighter oversight mechanism could indicate higher possibility of being found out or a higher importance of being transparent. These explanations are consistent with Proposition 1.

Using the first-order condition for Equation (19), we specify several testable comparative statics in Corollary 3.

Corollary 3: When the governance spending is sufficiently high for the donor to offer a positive performance-based sensitivity $w_1^* > 0$, the following statics hold:

i. $\frac{\partial b^*}{\partial \rho} < 0$, $\frac{\partial b^*}{\partial \lambda} < 0$, and $\frac{\partial b^*}{\partial c} < 0$

ii. $\frac{\partial b^*}{\partial \delta} > 0$ and $\frac{\partial b^*}{\partial r} < 0$, $\frac{\partial b^*}{\partial \sigma_\eta^2} < 0$

Part (i) suggests that NPOs with higher productivity, higher resources lost due to manipulation and higher reputation concerns will choose a lower level of misreporting. Part (ii) implies that the higher cost of exerting action is correlated with greater information misreporting; whereas higher risk aversion and higher organisation-specific risk variances induce a larger reporting accuracy. The intuition is that an NPO may opt to misrepresent financial information instead of spending higher costly effort to improve the terminal value. To see why the NPO would respond to a riskier environment (a higher r and σ_η^2) by reducing the extent of misreporting, notice from Equation (13) that all of the parameters in Corollary 3 have no direct effect on the misreporting their report, but through the effect on the incentive w_1^* as in Corollary 2. As the risk aversion and risk variances increase, the value-based incentive w_1^* is set lower to discourage the NPO from taking risk. The NPO now would have less incentive to manipulate the report due to the lower value-based incentive. The overarching intuition for Corollary 3 stems from the contract structure that dictates the compensation to be paid before the verifiable terminal value is realised. The agent faces a trade-off between expending efforts to improve the true state of the project and manipulate the report on which the contract is based.

Overall, we provide a simple model to parameterise the level of misreporting an NPO would commit if it is able to report the state of a funded project before the terminal

impact is realised. Under the assumption that the donor would form a belief about the NPO's misreporting strategy, we show that the level of optimal misreporting would depend on both the charitable effort (trade-off) and the oversight mechanism (cost of misreporting).

2.4. Econometric methodology

2.4.1. Empirical specification

Denote b_i the optimal degree of inaccuracy chosen by NPO i. We are primarily concerned with the effects of the choice of action (denoted a_i) and governance spending (denoted g_i) on the misreporting level. We capture the threshold effect of g_i on the effect a_i on b_i through an interaction term in the following specification:

$$b_i = \gamma_0 + \gamma_1 a_i + \gamma_2 g_i + \gamma_3 a_i \times g_i + \gamma_4 X_i + \varepsilon_i \tag{21}$$

where $\mathrm{I_i} = a_i \times g_i$ is the interaction between the optimal action a_i and the governance spending g_i; $\gamma_j, j \in \{1,2,3\}$ are the parameters of interest; X_i and γ_4 are respectively a vector of control variables and its vector of parameters; γ_0 represents a constant and ε_i is the error term.

The theory predicts that γ_1 and γ_2 should be positive; while γ_3 should be negative. These predictions are of a direct consequence of Proposition 1. Indeed, when g is set at $g = 0 < \tilde{g} = -\frac{\gamma_3}{\gamma_1}$, Proposition 1 suggests that $\gamma_1 = \frac{\partial b^*}{\partial a^*} > 0$. Likewise, when the NPO chooses zero effort, $a^* = 0 < \tilde{a} = \frac{\gamma_3}{\gamma_2}$, Proposition 2 suggests (Appendix 2.8.3) that $\gamma_2 = \frac{\partial b^*}{\partial g} > 0$. If g is a sufficiently high such that: $\frac{\partial b^*}{\partial a^*} = \gamma_1 + \gamma_3 g < 0$, we have $\gamma_3 < -\frac{\gamma_1}{g} < 0$.

Main variables of interest: For the dependent variable, we use the MAD statistic to measure the intensity of misreporting b_i. For non-binary explanatory variables, as we pool all of the available annual financial statements to construct the data in the main analysis, we take the averages of the respective financial figures over the years. First, we use the ratio of spending on charitable activities to the NPO's total income (*Charitable Spending)* to proxy for each NPO's choice of the optimal charitable action. Although the amount of income spent on charitable activities is not always a perfect signal to assess the charitable effort, it is highly correlated. Indeed, a non-profit with a higher ratio could be inferred to be exerting higher effort in maximising the use of their income, and this has long been used as the standard measure for the effectiveness of a charity (namely, program ratio reviewed in Hofmann and McSwain, 2013).

For governance spending, g_i, we use the proportion of the total income spent on governance activities. Table 2.A1 notes two types of governance activities: (1) spending on auditing and accounting activities, and (2) spending on administrative and other related activities. For the main analysis, we report the results when we lump the two types together as a sum (*Governance Spending)* since they could capture the disutility of misreporting as

discussed in the previous section. While a higher level of spending on administrative and accounting activities can reduce the level of genuine mistakes by improving the book-keeping process; higher levels of spending on administrative and auditing activities could increase the NPO's cost to manipulate financial statements creatively. Using either of the two expenditures yield qualitatively similar results.

The interaction term (*Interaction Term)*, I_i, is generated by multiplying *Charitable Spending* and *Governance Spending.*

Control variables: The set X_i aims to control for other observable characteristics and potential determinants that affect the precision of our measure of misreporting. We include the log of total assets (*NPO size*) to control for size; and the number of years the NPO has operated until the first survey (*Age)* to measure the NPO's establishment or familiarity with the sectoral norm (a standard practice, see Yetman and Yetman, 2011).

We also include the reported number of volunteers (*Volunteers*) to account for the fact that non-profit organisations are often overseen and run mainly by volunteers. In many cases, NPOs operate with modest internal accounting practices with volunteers serving as part-time bookkeepers (Keating and Frumkin, 2003). As the volunteers may receive little instructions or simply may not be fully committed, deficiencies in training and dedication can result in poor reporting accuracy. In contrast, having attracted a substantial base of volunteers could be a signal of the non-profit's strength of its philanthropy arms and concern about reputation. The consequence is the organisation becomes open and more transparent in their financial reports to maintain their position (see Corollary 3). The impact of volunteers depends on the balance between these two arguments. Although excluding this variable does not alter the core results, we discuss the importance of controlling for volunteers to our empirical strategy in the next section.

We include further six binary variables (Yes = 1, No = 0) that capture whether the NPO has ever: (1) reported expenditure on either internal or external audits (*Being audited)*, (2) received grants from any local, national or foreign government (*Receive government grants),* (3) reported zero fundraising expenses (*Zero fundraising*), (4) reported any losses from their investments/pension funds (*Losses from investments*), (5) received restricted income that is given for a specific purpose but within the charity's overall objective (*Receive restricted income*), and (6) received endowment funds (*Have endowment funds*). Since we pool all of the available annual financial statements, the binary variables equal to 1 if the corresponding variables take the value 1 at least once during the surveyed period, 0 otherwise. Previous non-profit studies indicate that the first three variables are expected to be associated with misreporting activities. Not having reports audited is a popular potential determinant of accounting fraud, probably associated with more irregularities in the data. Dependence on some specific types of donations, particularly from

governments, can lessen the non-profit's incentive to undertake illicit activities (see Garven et al., 2016). It also seems implausible that a non-profit could incur exactly zero expense in fundraising, hence reporting zero fundraising should infer some reporting inaccuracy. We add the last three control variables as potential determinants of misreporting. Incurring losses from investments or pension funds could induce the NPO's manager to manipulate their reports to hide the loss. Receiving restricted income and endowment funds could reduce the NPO's motivation for reporting manipulation: the charity upon receiving restricted income and endowment funds has ownership rights and will be acting as a principal instead of as an agent in the case of conduit giving. Despite the limited literature on this conjecture, we expect that as the charity has more power over their restricted income and greater reputation concerns for future receipts, they have the incentive to behave diligently and report accurate statements. They are also more likely to operate more sophisticated accounting systems that may be less prone to inaccuracy.[22]

To further test the impact of fundraising pressure and reputation concern outlined in Corollary 3 (part i), we include variable *Income from Donations/Grants*, constructed as the ratio of the income from charitable activities and voluntary sources, to capture the degree that a charity relies on incomes from donations and grants in addition to whether the charity relies on these sources.

Finally, we include the number of non-zero financial entries (*Number of non-zeros*) and the number of annual reports used to construct the measures (*Number of yearly reports*) to control for the size of the digit pool that potentially influences the precision of the measures. More non-zero figures could indicate a more diverse or complex NPO (for example, more activities), which could affect the degree of manipulation or human errors in the reporting process. By doing so, we also aim to control for NPOs who strategically withhold information by recording zeros. Once we control for the NPO's size and the number of yearly reports, more non-zero financial items being reported could be an

[22] A pitfall of using binary variables is the loss of information. Reducing the continuous variables to binary variables suits our context of potential misreporting for two reasons. First, the magnitudes of the reported continuous values could be manipulated (such as underreporting losses or over-reporting gains). As the binary variables are for the cumulated period (ever reporting losses for instance), NPOs would be more likely to underreport losses rather than record zero losses over the period. Binary variables could retain the information of, for example, whether the NPO experiences losses and such measures are less prone to measurement errors. Second, our results remain similar if we replace these last two binary variables by the ratios of restricted income/endowment funds to the total income (dependence on restricted income/ endowment funds) or exclude the variables from our specification.

indication of the NPO's openness. For that reason, we expect *Number of non-zeros* to have a negative effect on the extent of manipulation and errors.

2.4.2. Empirical strategy: traditional IV and Lewbel's (2012) approach

There are two concerns when estimating the effect of the charitable effort (measured by *Charitable Spending*) on the reporting behaviour of the NPO (measured by *MAD statistic).* First, there could be variables that affect both illicit behaviours and the organisation's tendency to exert effort. For example, a committed NPO would be likely to exert greater effort but be less likely to engage in manipulative activities. Second, the variable *Charitable Spending* could itself be mismeasured, and this measurement error increases the measure of misreporting *MAD statistic.* Since we expect a positive estimate of the effect (γ_1), a negative correlation between the unobserved commitment and the tendency to report less accurately would result in a downward attenuation bias. Our estimates would become closer to zero than the unbiased parameters, but the signs should remain. As such, measurement error is not a serious concern. We aim to address the omitted variable bias by: first excluding NPOs with unrealistic financial items (such as negative total assets or expenses); and second, including various control variables to mitigate the omitted variable bias. We extend the linear regressions by further employ two IV strategies to address the endogeneity concern, namely, the traditional method and Lewbel's heteroscedasticity IV estimation.

The traditional IV approach requires valid instruments that satisfy two criteria: being strongly correlated with *Charitable Spending* and *Interaction term* (strong identification), and orthogonal to the outcome variable after controlling for other potential confounders (exclusion restriction). Because endogeneity could also arise from measurement error in the outcome variable, the valid instruments also need to be excused from possible misreporting. Based on the data availability, we propose two instruments: the NPO's number of staff (*Headcount of staff*), and the actual spending on social security benefits (S*ocial Security spending*). These two instruments are less likely to be misreported by NPOs. The number of staff is easily either observed or cross-checked through employment contracts by the authority or the interested donors. Likewise, because the amount of social security cost is recorded in official/government papers, the NPOs will be restrained from falsifying these figures. The instruments are also likely to satisfy the other two criteria. For strong identification, we expect the number of staff and the amount of income spent on social security to positively correlate with charitable spending as more activities or services would require more paid employees, at least in the roles of supervision or planning. To account for the possibility of weak instrumentation, we also use the limited information maximum likelihood (LIML) procedure (Murray, 2006). For the exclusion restriction, because only a few specific staff such as the accounting division would have been involved

in misrepresenting financial information, it is difficult to argue that the employment size and social security contribution could have any direct impact on the misreporting behaviour. To our knowledge, there exists virtually no evidence to indicate the direct influence of the employment size on reporting accuracy and the incidence of accounting errors.[23] As we already control for the NPO size, spending on administrative/governance, auditing, and volunteering size, we expect that both the instruments can be excluded from the main Equation (21). Statistically, we report the Sargan-Hansen tests of over-identification to support the argument.

Another concern is that *Interaction term* could be endogenous due to the *Charitable Spending* component, although the Hausman endogeneity test fails to reject the equivalence of the estimates when we treat the *Interaction term* as exogenous. Nevertheless, we interact the proposed instruments with *Governance Spending* to construct two additional instruments and include into the estimations.

The second IV approach is proposed by Lewbel (2012). It allows for robustness checks of our traditional IVs when the standard exclusion restriction fails. The estimator exploits heteroscedasticity and higher moment conditions to construct internal instruments from the model's data without the need of any external source of variations. There are two conditions for identification. First, the error terms in the first stage regression are heteroscedastic. The greater the degree of heteroscedasticity in the error processes, the stronger will be the correlation of the generated instruments with the endogenous variables and the stronger first-stage prediction. We test the condition using the Breusch-Pagan test for heteroscedasticity in the first stage regressions. Second, there must exist a subset of the exogenous regressors uncorrelated with the covariance of the heteroscedastic error term and the second-stage error term. As discussed in Lewbel (2012), these conditions are normally satisfied in many models of endogeneity or mismeasurement, in which error correlations are due to some unobserved common factor. The misreporting context represents a valid setting as the main driving force of endogeneity discussed above is either the NPO's unobserved characteristics or mismeasurement error. There are two caveats of the estimator. First, using higher moment conditions is likely to provide less reliable estimates as it is not known how robust the results are to misspecification (Lewbel, 2012). Second, we are unable to acknowledge any economic intuition underlying the instruments. To address these concerns, we follow previous studies using the method (such as Emran

[23] Popular predictors related to employment are executive salaries (Keating et al., 2008), the size of the committee board, the presence of the audit committee (Krishnan and Yetman, 2011; see Garven et al., 2016 for a fuller discsussion in the non-profit literature). We are not able to include any of these variables due to data unavailability.

and Hou, 2013; Millimet and Roy, 2016; Loy et al., 2016; Caliendo et al., 2017), and supplement the set of internal instruments with our proposed instruments to improve the efficiency of the estimator. We briefly describe the estimator's intuition in Appendix 2.8.4.

Finally, the recorded spending on accounting and auditing services could also be subject to both measurement errors and confounders. We argue it is not a serious concern for three reasons. First, governance spending is often a part of the contract conditions externally set by funding bodies to warrant the transparency of charities (Hofmann and McSwain, 2016). Second, like *Charitable Spending,* the attenuation bias caused by mismeasurement would bias the estimates downward but not alter the signs of the estimate for *Governance Spending*. Third, the confounding effect of any unobserved commitment of the NPO would be mitigated by the various control variables, such as NPO sizes, volunteers, and whether the reports are audited. For completeness, we experiment treating *Governance Spending* as endogenous. As there is no reliable instrument for the variable, we use Lewbel's (2012) estimator to undertake the experiment and find qualitatively unchanged results. However, the respective Hausman test of endogeneity fails to reject the equivalence of treating the variable as exogenous at 5%. For that reason, we prefer treating *Governance Spending* as an exogenous variable in our main analysis.

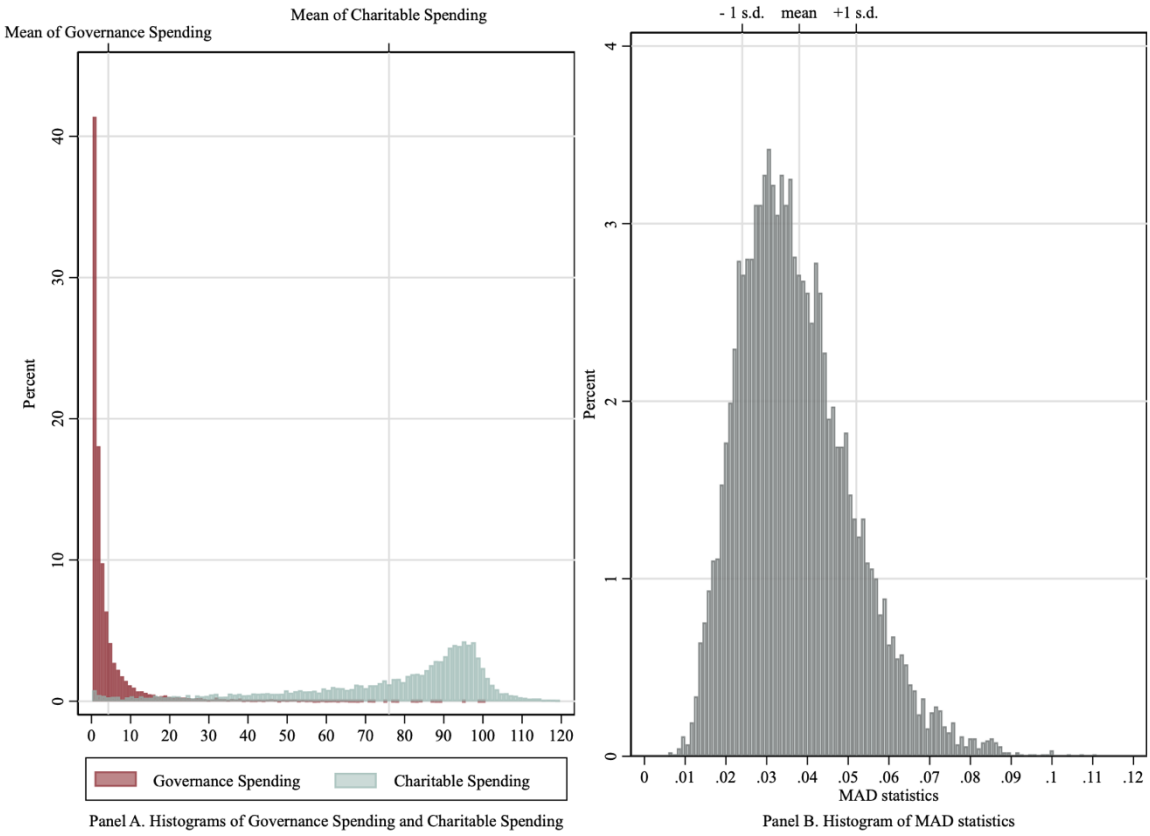
2.5. Empirical results

2.5.1. Descriptive statistics

Table 2.2 presents descriptive statistics for our variables and Table 2.3 compare the averages of the variables by conformity to Benford's Law for the sample used in the main analysis. As expected, conforming NPOs do have lower measures for deviations from the Benford distribution, as measured by MAD, χ^2, and KS statistics. Over the period of 2007 – 2015, on average, British NPOs spend 76% of their annual income on charitable activities, and 4% on governance activities. It is not surprising that the maximum values for the proportions of the total income spent on the two activities are large (nearly 800% and 100%, respectively). 16% of the sampled NGOs receive endowment funds and these NGOs could use their endowment funds to finance their activities. There is a clear difference in charitable spending between conforming and deviating NPOs. Conforming NPOs seem to spend more on charitable spending, be older, submitting more annual reports, but receive less government or restricted income and record more losses from investment. The figures for governance spending, size (total asset) and being externally audited are statistically indistinguishable between the two. There is also no difference in social security contribution and employment size (our proposed instruments) between the two types, reassuring our intuition that these variables should not directly correlate with the misreporting behaviour of the NGOs.

Figure 2.3 plots the histograms of three variables of interest: the MAD statistics, *Governance Spending* and *Charitable Spending.* Unsurprising, a large portion of the UK charities report a low percentage of total income spent on governance activities, echoing the survey results from Woodwell and Bartczak (2008) that 80% of the US non-profits lack financial resources for overhead and accounting costs. In contrast, a large portion of the UK charities report spending at least 80% of their total income on charitable activities. Panel B plots the distribution of the MAD statistics, while Figure 2.A2 in the Appendix 2.8. plots the statistics over five quantiles of key NPO characteristics such as charitable spending, governance spending, and size (total assets). Despite being descriptive, these correlations are consistent with our predictions in Section 3.

Figure 2.3. Histograms of the measures of misreporting and key expenses



Notes. Panel A plots histograms of *Governance Spending and Charitable Spending* as the percentages of the total income. Panel B the histogram of the MAD statistics. Number of bins is 100 for both panels. For presentation purposes, we exclude the top 1% of each variable from the histograms.

Table 2.2. Descriptive statistics of the main variables

VARIABLES	Mean	SD	Min	Max
MAD statistic	0.038	0.014	0.006	0.124
χ^2	35.44	25.17	1.328	326.2
KS	0.129	0.0623	0.0140	0.524
Charitable spending	0.763	0.315	0	7.797
Governance spending	0.043	0.092	0	1.01
Size (Total Assets, £ million)	9.551	97.27	6.93e-05	8,547
Age	20.87	14.80	0.564	50.89
Volunteers	9.717	249.8	0	17,500
Being audited	0.873	0.334	0	1
Receive government grants	0.384	0.486	0	1
Zero fundraising	0.542	0.498	0	1
Losses from investments	0.173	0.378	0	1
Receive restricted income	0.481	0.500	0	1
Have endowment funds	0.160	0.367	0	1
Income from Donations/Grants	0.816	0.252	0	1
Number of non-zeros	200.7	62.56	100	406
Number of yearly reports	5.626	1.285	2	8
Headcount	21.22	101.8	0	3,192
Social security spending (£)	77,294	361,013	0	1.517e+07

Note: The sample is restricted to 10322 British charities whose reports contain at least 100 non-zero financial items (transactions).

Table 2.3. Descriptive statistics by conformity to the Benford distribution

VARIABLES	All	Deviate	Conform	Difference	t-stat
MAD statistic	0.038	0.042	0.027	0.015***	52.282
χ^2	35.44	41.598	16.835	24.763***	47.747
KS	0.129	0.15	0.065	0.086***	75.347
Charitable spending	0.763	0.77	0.741	0.029***	4.000
Governance spending	0.043	0.044	0.042	0.002	0.404
Size (Total Assets, £ million)	9.551	8.894	11.537	-2.643	-1.193
Age	20.87	21.546	18.842	2.703***	8.046
Volunteers	9.717	7.271	17.108	-9.837*	-1.73
Being audited (binary)	0.873	0.871	0.877	-0.005	-0.702
Receive government grants (binary)	0.384	0.364	0.445	-0.081***	-7.366
Zero fundraising (binary)	0.542	0.543	0.54	0.002	0.214
Losses from investments (binary)	0.173	0.179	0.153	0.026***	3.07
Receive restricted income (binary)	0.481	0.474	0.504	-0.030***	-2.655
Have endowment funds (binary)	0.160	0.159	0.165	-0.006	-0.74
Income from Donations/Grants	0.816	0.812	0.828	-0.015***	-2.676
Number of non-zeros	200.7	202.592	195.183	7.409***	5.208
Number of yearly reports	5.626	5.696	5.414	0.282***	9.67
Headcount	21.22	21.437	20.558	0.879	0.379
Social security spending (£)	77.294	77599	76372	-1226.174	-0.149
Observations	10,322	2,567	7,555		

Notes: *** p <0.01, * p < 0.1. Observations are at the NPO level. The non-binary variables are averages of all NPO-year respective financial items over the period. The binary variables equal 1 if the respective variables take at least one non-zero value during the surveyed period. Conformity is based on Kolmogorov – Smirnov (KS) tests of the observed distribution following the expected distribution. At the 5% significance level, the subsample "Deviate" ("Conform") contains NPOs whose observed first digit distribution deviates from (conforms to) the Benford distribution. The reported t-statistics are for two-sided Wald tests on differences between the two subsamples' means.

2.5.2. Regression results

Table 2.3 shows estimates from an OLS with full control variables, the 2SLS with our proposed instruments, and the Lewbel's (2012) estimator.[24] To check how sensitive our instruments are to the specification (*Headcount,* and *Headcount* × *Governance Spending; Social Security spending,* and *Social Security spending* × *Governance Spending*), we first alternatively include either pair of instruments and test for the exogeneity of the *Interaction term* using the Wu-Hausman tests. If we fail to reject the null of statistical equivalence when treating *Interaction term* as being exogenous, we remove it from the set of endogenous variables in the sequential specification. We also experiment with treating *Governance Spending* as exogenous. Using the internal instruments generated by Lewbel's estimator, we test whether it is statistically equivalent when we do so. In Table 2.A3 we fail to reject the equivalence when treating *Governance Spending* as exogenous. This evidence further supports the theoretical assumption in the literature and our premise that *Governance Spending* should be treated as externally determined by external bodies (such as due to resources constraints). Our preferred specification is a 2SLS with both pairs (and their interactions with *Governance Spending)* as the instruments in Column (3) for *Charitable Spending* and *Interaction term* and treating *Governance Spending* as exogenous. To improve the efficiency, our preferred specification of the Lewbel's estimation is estimated using GMM (Baum et al., 2012). To check the robustness of our tradition instrumentation, we exclude the external instruments and rely on internally generated instruments from the Lewbel's estimator. We report a range of similar results from other specifications when we treat *Governance Spending* as endogenous in Table 2.A3.

Once we control for a battery of organisational characteristics or using the two IV strategies, we find robust evidence for the thresholds laid out in the theoretical predictions. The marginal effects of both *Charitable Spending* and *Governance Spending* on the measure of misreporting (MAD) are non-monotonic but instead dependent on the magnitudes (the thresholds) of the other expense. To determine the thresholds in our dataset for British charities, we derive the following thresholds from Equation (21).

$$\tilde{a} = -\frac{\text{Coefficient of } \textit{Governance spending}}{\text{Coefficient of } \textit{Interaction term}} \tag{22}$$

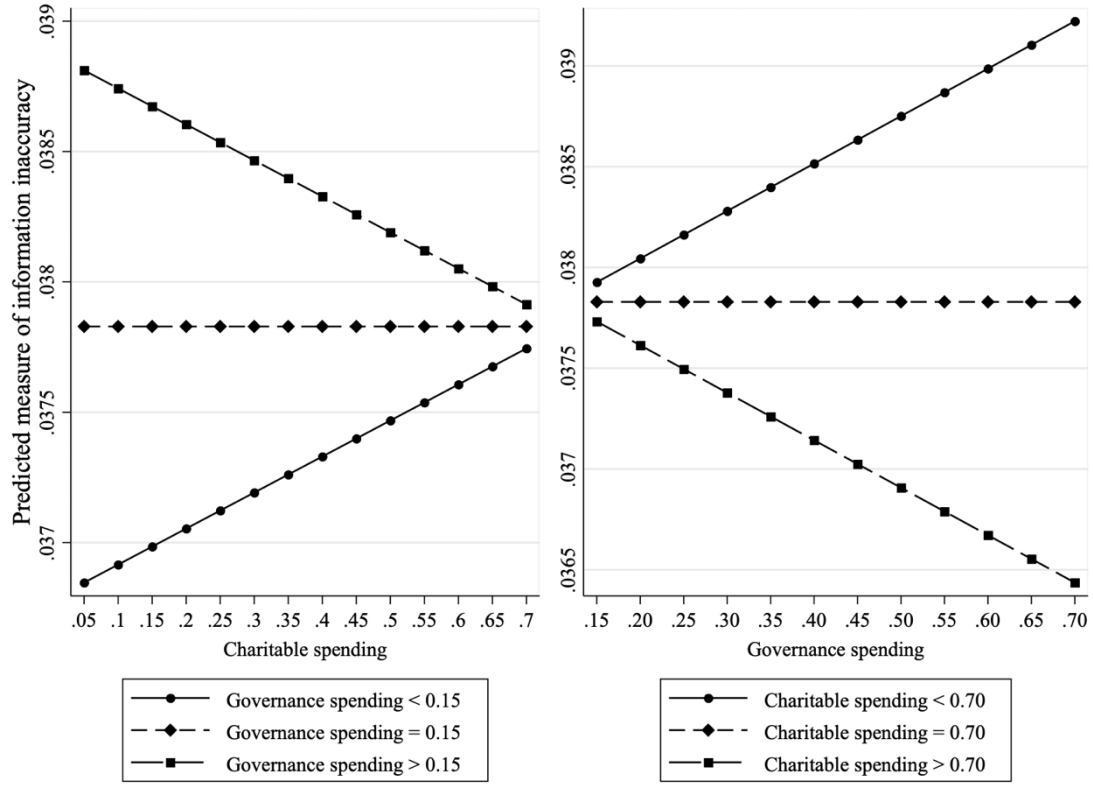
$$\tilde{g} = -\frac{\text{Coefficient of } \textit{Charitable Spending}}{\text{Coefficient of } \textit{Interaction term}} \tag{23}$$

Back-of-the-envelope calculations suggest using the OLS with control variables and the Lewbel's estimation (our preferred specification) suggest that $\tilde{a} \approx 70\%$ and $\tilde{g} \approx 15\%$. That

[24] We use Variance Inflation Factor analysis (VIF) to see if multicollinearity drives our results. Small condition indices (substantially lower than 10) indicates it is not the case.

is, only organisations who exert relatively high effort in charitable activities (spending 70% of their total income), their higher spending on governance spending would resonate with a lower measure of misreporting. Otherwise, higher governance spending needs not translate to a higher level of reporting accuracy. Similarly, unless charities spend sufficiently on auditing and accounting activities (about 15% of total income), charities performing well (spending more on charitable activities) would be associated with a higher level of reporting accuracy. To put the number in perspectives, only more than 7% of the charities in our sample spend at least 15% of their total income on governance activities. Despite being suggestive, the 15% figure highlights the lack of resources spent on governance activities in the UK sector relative to our threshold. The result supports the call for increased support from granting donors on overhead costs and accounting expenses as documented by Singh (2015). Figure 2.4 visualises these marginal effects and thresholds.

Figure 2.4. The predictive marginal effects on inaccuracy of financial reports



Notes: Margin plots of the marginal effects obtained after our preferred Lewbel (2012) estimation. The left (right) panel shows marginal effect of charitable spending (governance spending) on the predicted measures of misreporting. Lines with cirles (squares, diammonds) represent the respective marginal effect when the value of the variable on the horizontal axis is below (above, at) the threshold (0.15 for governance spending, 0.70 for charitable spending). *Sources:* Authors' calculation using the UK Third Sector Research Data Collection.

Table 2.4 also provides interesting estimates for other organisational characteristics. Contradictory to as Krishnan et al. (2006), size and age are positively correlated with the amount of information irregularities. The result is hardly surprising in the non-profit literature. For example, Keating and Erumkin (2003) suggest larger non-profits with manual accounting systems may be prone to errors if they do not change the system to adapt to the loading tasks. Having reports audited, receiving government grants, restricted income or endowment funds are significantly correlated to lower levels of irregularities. Reporting zero fundraising and losses from investments/pension funds do not significantly correlate with worse reporting accuracy, despite hypothetical motivations for the NPO to falsify their data. NPOs receiving more incomes from the public or grants, however, would report more accurately. This result is consistent with the explanation of reputation concern – those who would like to remain trustworthy to the public are more likely to well behave. Having more non-zero financial items recorded, controlling for the number of yearly reports, is significantly correlated with better accuracy. One reason could be that NPOs disclosing more non-zeros are indeed those reporting truthfully. We further test this concern in Appendix 2.8.6.4.

Table 2.5 supports the statistical validity of our instruments. The Kleibergen-Paap rk Wald F-statistics (to account for potential heteroskedasticity and two potential endogenous variables) are large in all cases, supporting the relevance of our traditional instruments and the instruments generated by Lewbel's estimator. We also reject the null of homoscedastic errors in the first-stage estimations, satisfying the first identification condition of the Lewbel's (2012) estimator. We fail to reject the null that the traditional instrumentation is not overidentified using the Hansen J statistics. To test for the overidentification of the Lewbel's (2012), we also report C-statistic to test for the orthogonality of suspect instruments (see Hayashi, 2000, pp. 227-8). C-statistics provides additional information over Hansen J statistics for two aspects. First, for a model with a large number of instruments, a Hansen-Sargan test may have little power (Baum et al., 2003). Second, C-statistics allows us to test for the statistical validity of suspect instruments, that is, the instruments that we deem to have statistical validity. Since we wish to test for the robustness of our results using the two traditional instruments, C-statistics allows us to exclude the overidentification conditions associated with the traditional instruments, only testing the statistical validity of the internally generated instruments. Although the Hansen J statistics in Table 2.5 rejects the null of overidentification, C-statistics supports the statistical validity of the internally generated instruments. As a cautionary note, our tests of overidentification cannot test for the excludability assumption, but instead the coherency of the instruments, that is, whether the instruments identify the same parameters (Parente and Santos Silva, 2012).

Table 2.4. Main results for determinants of misreporting for British charities

VARIABLES	*Dependent variable:* MAD statistics			
	(1) OLS	(2) OLS_with controls	(3) 2SLS	(4) Lewbel (2012)
Charitable spending	0.300	3.362***	1.981**	2.359***
	(0.547)	(0.589)	(0.836)	(0.731)
Interaction term	-7.848	-16.251***	-10.613***	-16.111***
	(5.303)	(5.513)	(3.364)	(5.123)
Governance spending	13.585***	11.957***	6.298***	11.279***
	(2.975)	(3.130)	(1.934)	(3.120)
Size		0.515***	0.925***	0.525***
(logged total assets)		(0.083)	(0.237)	(0.081)
Age		0.080***	0.055***	0.079***
		(0.009)	(0.016)	(0.009)
Volunteers		-0.208	-0.146	-0.151
(number of volunteers)		(0.319)	(0.253)	(0.311)
Being audited		-2.483***	-2.509***	-2.456***
(1 = Yes, 0 = No)		(0.463)	(0.495)	(0.461)
Receive government grants		-1.974***	-1.910***	-1.961***
(1 = Yes, 0 = No)		(0.263)	(0.278)	(0.262)
Zero fundraising		-0.189	0.819	-0.220
(1 = Yes, 0 = No)		(0.271)	(0.666)	(0.270)
Losses from investments		0.354	0.416	0.345
(1 = Yes, 0 = No)		(0.347)	(0.372)	(0.345)
Receive restricted income		-0.543*	-0.544*	-0.519*
(1 = Yes, 0 = No)		(0.292)	(0.312)	(0.291)
Have endowment funds		-1.102***	-0.635	-1.187***
(1 = Yes, 0 = No)		(0.345)	(0.466)	(0.345)
Income from Donations/Grants		-1.583***	-5.644**	-1.288**
		(0.577)	(2.219)	(0.574)
Number of non-zeros		-0.170***	-0.193***	-0.169***
		(0.005)	(0.014)	(0.005)
Number of yearly reports		4.069***	4.883***	4.026***
		(0.209)	(0.513)	(0.209)
Constant	37.313***	42.401***	27.689***	42.922***
	(0.449)	(1.520)	(7.933)	(1.542)
R-squared	0.004	0.263	0.174	0.262
Observations	10,322	10,322	10,322	10,322

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Column (1) and (2) are baseline results. Column (3) use *Headcount,* and *Headcount* × *Governance Spending; Social Security spending,* and *Social Security spending* × *Governance Spending* as instruments for *Charitable Spending* and *Interaction Term,* treating *Governance Spending* as exogenous. Columns (4) reports our preferred estimates from the Lewbel's estimator. Estimates are scaled up by 1000 for ease of interpretation.

Table 2.5. Diagnostic tests for the IV estimations

Panel A. Diagnostic tests for the main estimation		
	Estimators	
	2SLS	Lewbel's (2012)
(underidentification) Kleibergen-Paap rk LM stat	54.461	306.401
(p-value)	(0.000)	(0.000)
(weak identification) Kleibergen-Paap rk Wald F-statistics	18.682	128.623
(overidentification) Hansen J statistics for 2SLS/C-statistics for Lewbel's (2012) (p-value)	0.217	52.761/6.800
	(0.641)	(0.001)/(0.147)
(endogeneity) Wu-Husman test of endogeneity	5.466	6.575
Chi-square(1) (p-value)	(0.019)	(0.010)
Panel B. Diagnostics for first-stage estimations of the Lewbel's estimators		
First-stage estimation for *Charitable spending*		
Breusch-Pagan / Cook-Weisberg statistics for heteroskedastic errors - Chi-square (1) (p-value)		47.35***
		(0.000)
First-stage estimation for *Interaction term*		
Breusch-Pagan / Cook-Weisberg statistics for heteroskedastic errors - Chi-square (1) (p-value)		620.52***
		(0.000)

Notes: Kleibergen-Paap rk Wald F-statistics are under the null that instruments are weak for iid being violated. Wu-Hausman tests are for equivalence of the estimates under exogeneity. Breusch-Pagan / Cook-Weisberg LR tests are under the null H_0 that the errors are homoscedastic.

2.6. Sensitivity Analyses

In this section, we present our preferred robustness check using within-NPO variations to compliment the between-NGO comparison in the main analysis. Since the data span over an eight-year period, it allows us to separate the sample into two periods (2008 – 2011 and 2012 – 2015) while maintaining the number of non-zeros (100) required to construct our measures of misreporting based on Benford's Law. The procedure is as follows. We split the sample into two parts by two four-year intervals and re-calculate the measures of misreporting and the explanatory variables as described in Section 2.4.1 for each part. We then match the two parts and obtain a panel dataset of 4318 NPO-year observations (2159 NPOs that have at least 100 non-zero financial points in each period).

Controlling for within-NPO time-invariant and time-varying NPO characteristics (omitted to save space), we obtain similar results and thresholds as in our main analysis. Columns (2) and (3) in Table 2.6 suggest British charities seem to misreport more intensively over time and the temporal difference is statistically significant. One potential explanation is that because there are more opportunities for funding in the post-crisis period 2012 to 2015, NPOs are more inclined to produce more favourable financial reports to attract the funding.

Table 2.6. Fixed effect models for misreporting over two periods

VARIABLES	*Dependent variable:* MAD		
	(1) OLS	(2) OLS	(3) OLS
Charitable spending	3.615**	3.100*	3.377**
	(1.822)	(1.699)	(1.731)
Interaction term	-13.557***	-8.505*	-9.405**
	(4.793)	(4.422)	(4.445)
Governance spending	7.916	6.479	7.143
	(6.351)	(5.287)	(5.294)
Year Dummy (Period 2 = 1)		2.919***	2.333***
		(0.287)	(0.744)
Constant	35.092***	34.829***	19.786*
	(1.480)	(1.375)	(10.307)
NPO-Year observations	4318	4318	4318
Number of NPOs	2159	2159	2159
Control Variables	No	No	Yes
NPO fixed effect	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. To obtain the sample, we split the original sample into two periods 2008 – 2011 and 2012 – 2015. We pool four years of data in each sample and re-calculate measures of misreporting and explanatory variables as in Section 2.4.1. Control variables are listed in Table 2.3 or Section 2.3.1. Estimates are scaled up by 1000 for ease of interpretation.

We perform further four sensitivity checks in Appendix 2.8.6. While the precise magnitudes of the estimates of interest vary depending on the sub-samples, the major conclusions remain. Specifically, our results are not sensitive to the constructing algorithms of our measures; the cut-offs of non-zeros we use in the construction of the measures; extreme values (outliers) of NPO sizes, spending on charitable or governance activities; and finally, potential selection bias from non-disclosure as we construct the measures for organisations that have at least 100 financial figures after pooling yearly data. Overall, we demonstrate the robustness of our results.

2.7. Concluding remarks

We provide the first systematic study on the reporting behaviour of non-profits. We advocate the use of Benford's Law as an alternative to measuring irregularities in financial data. We find financial figures from 25% of the charities collected in the UK Third Sector Data during the period 2008-2015 do not conform to Benford's Law at the 5% significance level, suggesting potential irregularities. The approach is a computationally easy and useful screening step to identify potential organisations for an extensive investigation, but we emphasise that it does not provide definitive evidence of fraudulent behaviour, nor does it substitute auditing. Instead, we view our method as a way to improve the efficiency of

assessing charities' financial datasets, reduce the costs of monitoring the sector, and put pressure on non-profits to conform with the raising accountability norm.

We also support a leading voice from the charity sector (Singh, 2015) that over-spending on governance activities and back-offices could be counter-productive by failing to motivate the organisations to adhere to accountability. NPOs' charitable effort and the oversight mechanisms, measured by spending on charitable activities and governance activities respectively, need to be sufficiently high (70% and 15% of total income) so that their public financial data become credible. Unless funders and the public consider support for both the activities in their funding package, the accountability pressure could distort philanthropic agendas of the NPOs. Although our thresholds should be interpreted as indicative due to non-experimental data, our work provides the first step to identify relevant indicators to assist regulators or donors when assessing programme ratios and giving support package to overhead activities. In addition to evaluating NPOs performance on self-reported information, donors could extend other methods such as client-based assessments and surveys. We hope to open further research and discussions on these issues.

2.8. Appendix For Chapter 2

Table 2.A1. Financial items reported by the NPOs

Panel A.	Panel B.
1. ASSETS	**5. INVESTMENTS**
Net Assets	Investments
Net Current Assets	Investment – Rent from property
Current Assets	Investment – Dividends
Cash in hand or at the bank	Investments – Interest on deposits
Debtors	
Current Investments	**6. OTHER INCOMES**
Stocks	Voluntary income– Government Sector
Creditors due within one year	Voluntary income– Central Government
Pension Assets	Voluntary income– Local Government
Fixed Assets	Voluntary income– Regional Government
Intangible Fixed Assets	Voluntary income– Town and Parish Councils
Investment Assets	Voluntary income– NHS Trusts
Tangible Fixed Assets	Voluntary income– European Government
Creditors due after one year	Voluntary income– International Government Agency
Other Assets	Voluntary income– Foreign Governments
Provisions	Voluntary income– Public Corporations
	Voluntary income– Universities
	Voluntary income– Devolved Governments
2. EXPENDITURES	Voluntary income– Business Sector
Expenditures	Voluntary income– Nonprofit sector
Expenditures on Charitable Activities	Voluntary income– General public
Expenditures on generating funds	Voluntary income– Government Sector
Expenditures on fundraising and publicity	
Expenditures on investment management	**7. STAFF**
Expenditure on trading subsidiary	Number of Full-time staff (FTE)
Costs of obtaining voluntary income	Number of visitors
Costs of processing grants	Number of other non-stipendiary participants
Total costs of governance *	Staff headcount
Costs of Accounting and Audit Fees	Number of Volunteers
Costs of Administrative	Number of Audit and Accounting staff
Other Governance Expenditure	
3. FUNDS	**8. OTHER INFORMATION**
Total Funds	Depreciation (value)
Endowment Funds	Endowment received (value)
Income Funds	Revaluations of fixed assets
Restricted Funds	Gains/Losses on Investments
Unrestricted Funds	Interest payments
Other Funds	Income by fund *
Pension Funds	Income from endowment
4. INCOME	Other income

Incoming resources (Total)*	Restricted income
Charitable activities – Government Sector	Unrestricted income
Charitable activities – Central Government	Other financial values
Charitable activities – Local Government	Gains/Losses on pension funds
Charitable activities – Regional Government	Reserves
Charitable activities – Town and Parish Councils	Staff costs
Charitable activities – NHS Trusts	Other staff costs
Charitable activities – European Government	Pension costs
Charitable activities – International Government Agency	Social security costs
Charitable activities – Foreign Governments	Wages and salaries
Charitable activities – Public Corporations	Support costs
Charitable activities – Universities	Irrecoverable VAT
Charitable activities – Devolved Governments	Tangible fixed assets
Charitable activities – Business Sector	Additions
Charitable activities – Nonprofit sector	Net book value – beginning
Charitable activities – General public	Disposals
Charitable activities – Government Sector	Net book value – end
Income from Funds (Total)*	
Generating funds – Government Sector	
Generating funds – Central Government	
Generating funds – Local Government	
Generating funds – Regional Government	
Generating funds – Town and Parish Councils	
Generating funds – NHS Trusts	
Generating funds – European Government	
Generating funds – International Government Agency	
Generating funds – Foreign Governments	
Generating funds – Public Corporations	
Generating funds – Universities	
Generating funds – Devolved Governments	
Generating funds – Business Sector	
Generating funds – Nonprofit sector	
Generating funds – General public	
Generating funds – General public	

Notes: * indicates items not included in the Benford's Law digital analysis due to duplicate. All financial items are from the surveys and recorded in Alcock, & Mohan (2015) in Sterling (£). For purpose of the study, we convert non-Sterling figures to the contemporary values in Sterling (£). Because of the scale invariability property, this does not affect the applicability of Benford's Law.

2.8.1. Proof of Lemma 1

Given $\omega = w_0 + w_1(\rho a + \eta + b - b^e - \lambda b^e) - \frac{\delta}{2}a^2 - \frac{g}{2}(b-\tau)^2 - c(b-b^e)$, the induced wealth is normally distributed $\omega = \mathrm{N}(\mu, d^2)$ with $\mu = w_0 + w_1(\rho a + b - b^e - \lambda b^e) - \frac{\delta}{2}a^2 - \frac{g}{2}(b-\tau)^2 - c(b-b^e)$ and $d^2 = w_1^2\sigma_\eta^2$. We can show that:

$$E[u(\omega)] = \int_{-\infty}^{+\infty} \frac{1}{d\sqrt{2\pi}} \left(-r\omega - \frac{(\omega-\mu)^2}{2d^2}\right) \tag{A1}$$

$$-r\omega - \frac{(\omega-\mu)^2}{2d^2} = -\frac{1}{2d^2}[(\omega-\mu) + rd^2]^2 - r\mu + \frac{r^2d^2}{2} \tag{A2}$$

(A1) and (A2) imply that:

$$E[u(\omega)] = \exp\left(-r\omega + \frac{r^2d^2}{2}\right) \int_{-\infty}^{+\infty} \frac{1}{d\sqrt{2\pi}} \exp -\frac{1}{2d^2}[(\omega-\mu) + rd^2]^2 d\omega \tag{A3}$$

that leads to $E[u(\omega)] = -\exp(r\mu + \frac{r^2d^2}{2}) = -\exp(r\mathrm{CE})$.

Or $\mathrm{CE} = \mu - \frac{rd^2}{2}$. The proof completes.

2.8.2. Proof of Proposition 1

We first notice that:

$$\left.\frac{\partial b^*}{\partial a^*}\right|_{\partial g} = \left.\frac{\partial b^*}{\partial g}\right/\frac{\partial a^*}{\partial g} \tag{A4}$$

Because $a^* = \frac{\rho}{\delta}\left(\frac{\frac{\rho^2}{\delta} - \frac{\lambda}{g}}{\frac{\rho^2}{\delta} + \frac{1}{g} + r\sigma_\eta^2}\right) = \frac{\rho}{\delta}\left(\frac{(1+\lambda)\frac{\rho^2}{\delta} + \lambda r\sigma_\eta^2}{\frac{\rho^2}{\delta} + \frac{1}{g} + r\sigma_\eta^2} - \lambda\right)$, we have:

$$\frac{\partial a^*}{\partial g} > 0, \qquad \forall \mathrm{g} \geq 0 \tag{A5}$$

From Equation (14):

$$b^* = \frac{\beta - \frac{\lambda}{g}}{g\beta + g\Delta + 1} - \frac{c}{g} + \tau \tag{A6}$$

Differentiating b^* with respect to g we have:

$$\text{sign } \frac{\partial b^*}{\partial g} = \text{sign } T(g) \tag{A7}$$

where: $T(g) = -[\beta - c(\beta+\Delta)]g^2 + (\lambda + c)g + \frac{\lambda + c}{\beta + \Delta}$.

We examine sign $T(g)$ with respect to g. $T(g)$ has two roots as of:

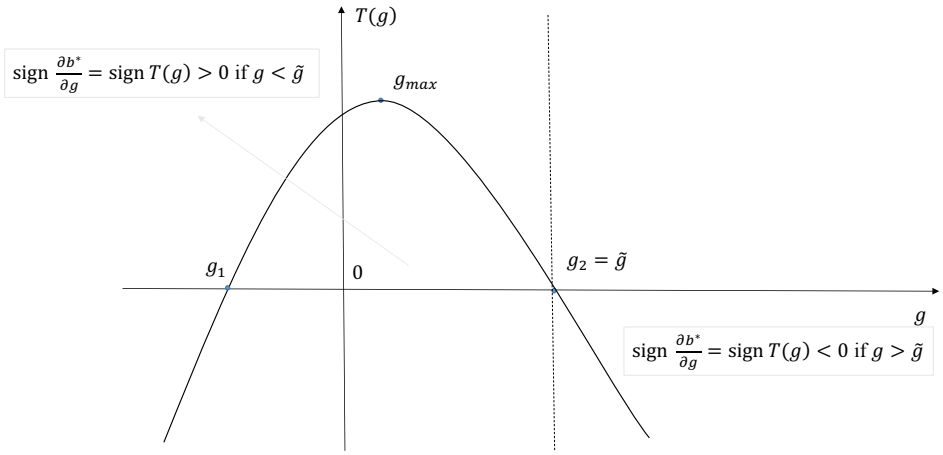
$$g_{1,2} = \frac{\lambda + c \mp \sqrt{(\lambda+c)\lambda + \frac{\beta(\lambda+c)}{\beta+\Delta}}}{\beta - c(\beta+\Delta)} \tag{A8}$$

and the maximal point at: $g_{max} = \frac{\lambda+c}{2(\beta - c(\beta+\Delta))} > 0$

As we assume $\beta > c(\beta + \delta)$, following Descartes' rule of signs we have $g_1 < 0 < g_2$. To save space, we provide a graphical proof for ease of interpretation. Figure 2.A1 shows that, given $g \geq 0$ sign $T(g) < 0$ if and only if $g \geq g_2$ and sign $T(g) > 0$ iff $g < g_2$.

Following (A6), sign $\frac{\partial b^*}{\partial g} < 0$ if and only if $g \geq g_2$ and sign $\frac{\partial b^*}{\partial g} > 0$ if and only if $g < g_2$. Combining with (A5), $\frac{\partial b^*}{\partial a^*}\Big|\partial g = \Big/\frac{\frac{\partial b^*}{\partial g}}{\frac{\partial a^*}{\partial g}} < 0$ if and only if $g \geq g_2$ and $\frac{\partial b^*}{\partial a^*}\Big|\partial g > 0$ if and only if $g < g_2$. Set g_2=$\tilde{g}$ and the proof completes.

Figure 2.A1. How sign $\frac{\partial b^*}{\partial g}$ and sign T(g) behave when g varies in $(0, +\infty)$.



Source: Authors' own illustration.

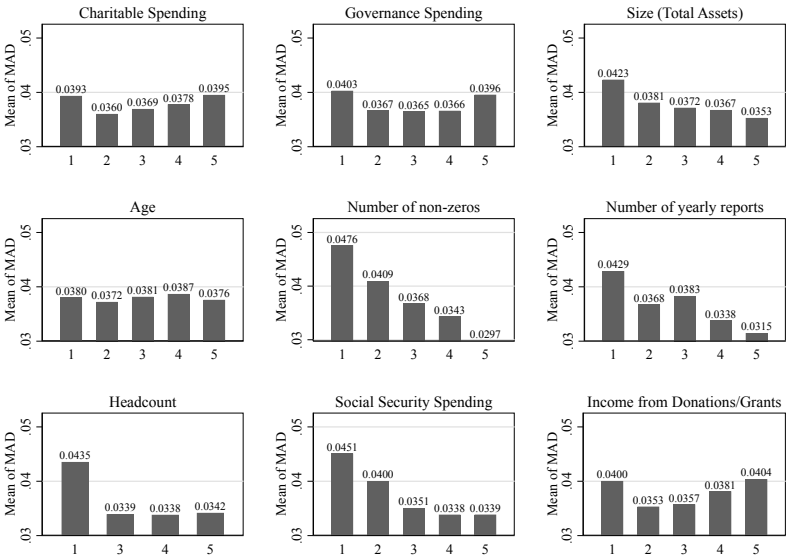
2.8.3. Proof of Proposition 2

Proof: As $\frac{\partial a^*}{\partial g} > 0$ $\forall$g ≥ 0, $g \geq \tilde{g}$ if and only if $a \geq a_g = \tilde{a}$ and $g < \tilde{g}$ if and only if $a < a_g = \tilde{a}$ with $\tilde{g} = g_2$ specified as in Figure 2.A1. Figure 2.A1 also confirms that $\frac{\partial b^*}{\partial g} < 0$ if and only if $g \geq \tilde{g}$; we immediately have that $\frac{\partial b^*}{\partial g} < 0$ if and only if $a \geq \tilde{a}$.

The proof completes.

2.8.4. Additional summary statistics

Figure 2.A2. Measures of misreporting (MAD) by quantiles of NPO characteristics



Notes: MAD statistics is plotted over five quantiles of nine NPO characteristics.

2.8.5. Lewbel' (2012) IV estimator

Assume that the model of interest is: $Y_1 = X'\beta_1 + \gamma_1 Y_2 + \varepsilon_1$ and the endogeneity problem of Y_2 emerges from $Y_2 = X'\beta_2 + \varepsilon_2$ (B.7), where X is a set of exogenous regressors. The traditional IV approach assumes that some elements of vector X are non-zero in (19) (strong identification) but zero in (18) (exclusion restriction). Lewbel's theorem shows that the parameters are identified if there exist exogenous variables $Z \subseteq X$ and heteroscedasticity in the data such that $E(Z'\varepsilon_1) = E(Z'\varepsilon_2) = cov(Z, \varepsilon_1\varepsilon_2) = 0$ and $cov(Z, \varepsilon_2^2) \neq 0$. The variables $(Z - E(Z))\varepsilon_2$ can then be used as instruments for Y_2. Lewbel proves that the assumptions $cov(Z, \varepsilon_2^2) \neq 0$ and $cov(Z, \varepsilon_1\varepsilon_2) = 0$ are analogous to the two criteria under the traditional IV approach and they ensure $(Z - E(Z))\varepsilon_2$ to be a valid instrumentation. In our context, assuming both *Charitable spending* and *Interaction term* are endogenous, the estimator can be implemented as follows:

i. Regress *Charitable spending* on the set X_i by OLS and save the residuals, ε_{1i}.

ii. Regress *Interaction term* on the set X_i by OLS and save the residuals, ε_{2i}.

iii. Form instruments $Z_{ij} = (X_i - \overline{X_i})\varepsilon_{ji}$ with $j = 1,2$

iv. Estimate the main equation of interest (16) via GMM using $Z_{ij}, j = 1,2$ as instruments for *Charitable spending* and *Interaction Term.* GMM is preferred to 2SLS as the set of exogenous variables X_i contains more than one element, 2SLS becomes prone to over-identification and should be efficiently estimated with GMM (Baum et al., 2003).

v. Add the traditional instruments to improve the efficiency and avoid overidentification.

2.8.6. Robustness checks

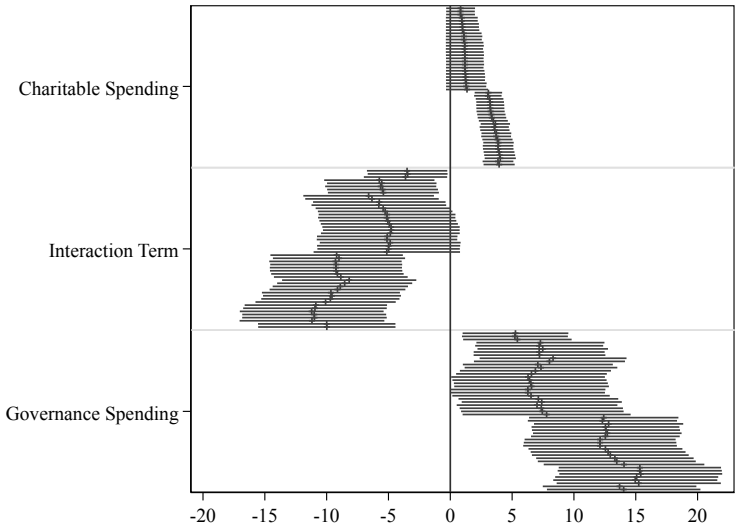
2.8.6.1. Sensitivity to the measurement of reporting inaccuracy

We replace the MAD statistics by the three critical-based alternatives: χ^2 (Chi-square statistics), KS statistics, and the binary variable *Deviate* indicating whether the non-profit's data deviate from the Benford distribution. Section 2. specifies how these measures are constructed. Table 2.A3 reports the estimates from OLS with full controls, our preferred 2SLS (2SLS-6 in previous tables), and Lewbel's (2012) estimator. Although we cannot directly compare the magnitudes of the coefficients, all the signs and significance are unchanged, supporting our results' robustness.

2.8.6.2. Sensitivity to the cut-off of the number of non-zero items

One concern when using Benford's Law in a digital analysis is the cut-off of the number of non-zero financial observations to include in the data pool. In the main analysis, we use the rule-of-thumb threshold of 100. We explore how our results are sensitive to the cut-off. We also address the concern that some non-profits may have withheld some information by recording zero financial transactions. If the mechanisms underlying the decision to withholding information and manipulating the reported information are similar, we should not observe any systematically different results when we include NPOs with more zero financial items, who are less likely to withhold information. We vary the cut-off from 115 to 65 and re-do the analysis 50 times. The unreported results are quantitatively unchanged, only the estimates become less precise when the thresholds fall below 75. We conclude that the cut-off choice does not drive the results.

Figure 2.A3. Robustness to varying the cut-offs of non-zero financial points

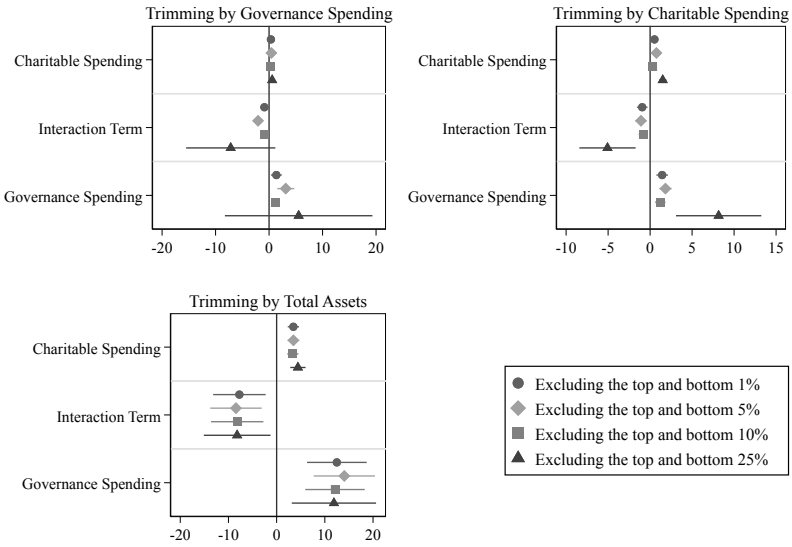


Notes: The estimates of the three variables of interest are in a descending order of the number of non-zeros used from 65 to 115. For example, the top estimate is for Charitable Spending at the cut-off of 65. Confidence level is at 95%. The graph shows a clear robustness to our main results when varying the cut-offs.

2.8.6.3. Sensitivity to the sample used

As the distributions of total assets and spending in the UK third sector are heavily skewed, there are two concerns. First, our results may be driven by outliers. Second, the skewness could introduce heteroscedasticity to our linear estimation. Using the specifications in the main analysis, we perform various trimming exercises: alternatively excluding the top and (or) the bottom 1%, 5%, 10%, 25% percentile of the *Size* (total assets), *Governance Spending,* and *Charitable Spending*. Figure 2.A4 summarise the main results from this check. Overall, the magnitudes and the significance of the main estimates are robust to trimming off outliers and potential heteroscedasticity.

Figure 2.A4. Robustness to excluding various ranges of outliers



Note. The checks use OLS with full controls as in the main analysis.

2.8.6.4. Controlling for potential informational non-disclosure

Because we exclude NPOs with fewer than 100 non-zero financial items, there may be an issue of selection. The excluded NPOs could either have operated in more straightforward/fewer activities which generate no significant transactions or have strategically withheld information by recording some significant items as zero. Although our main results are not sensitive to the threshold of 100, we use the Heckman sample correction to add further evidence in support of our findings' robustness.

Let T be a binary variable taking value 1 if the NPO reports at least 100 non-zeros in our sample and 0 otherwise. In the first stage, we explore of the selection of NPOs to record more than 100 non-zero financial items by running a probit regression of T on explanatory variables X_i. In the second stage, we include the inverse Mills ratio obtained from of the predicted individual probabilities in the first stage as additional explanatory variables. Table 2.A2 presents the estimates for the two stages. Even when controlling for

potential selection bias due to excluding NPOs who record fewer than 100 non-zeros, our results remain qualitatively unchanged for all the four indices. We report Wald tests of independence under the null that the two decisions can be taken independently. Although we reject the null for the MAD statistics, we fail to do so for the other critical-based measures. The identification for this model is based on the normality assumption when the same covariates appear in both the two-stage equations. Despite being tenuous, we note that having losses from investment and reporting zero fundraising costs are significant determinants of providing more than 100 financial figures; while they are always insignificant in explaining the reporting inaccuracy. As such, these two variables can work as the exclusion restriction controls for our Heckman's correction model.

Table 2.A2. Heckit estimator for missing observations for the four indices

VARIABLES	First Stage	Second stage (degree of misreporting)			
		MAD	χ^2	KS statistics	Deviate
Charitable spending	0.13**	3.16***	7.37***	17.07***	112.55***
	(0.53)	(0.56)	(1.19)	(2.56)	(18.12)
Interaction term	-0.12	-7.78***	-12.77**	-38.50***	-185.67**
	(0.21)	(2.76)	(5.08)	(11.84)	(72.30)
Governance pressure	0.26**	5.68***	10.92***	31.27***	163.69***
	(0.12)	(1.57)	(2.75)	(7.41)	(41.64)
Size	0.27***	0.61***	1.60***	2.42***	18.20***
(logged total assets)	(0.14)	(0.08)	(0.18)	(0.39)	(2.90)
Age	-0.048***	0.08***	0.16***	0.30***	1.18***
	(0.02)	(0.01)	(0.02)	(0.04)	(0.31)
Volunteers	10.27*	-0.22	-0.45	-1.66*	-28.44**
	(6.10)	(0.31)	(0.32)	(0.86)	(12.65)
Being audited	8.40***	-2.37***	-2.81***	-9.29***	-34.24**
(1 = Yes, 0 = No)	(0.51)	(0.47)	(0.92)	(2.20)	(14.25)
Receive government grants	6.84***	-1.94***	-3.29***	-7.75***	-43.50***
(1 = Yes, 0 = No)	(0.54)	(0.26)	(0.55)	(1.27)	(9.67)
Zero fundraising (1 = Yes, 0 =	5.08***	-0.04	0.14	-0.65	13.11
No)	(0.48)	(0.27)	(0.56)	(1.29)	(9.46)
Losses from investments	4.91***	0.32	-0.25	0.71	-1.84
(1 = Yes, 0 = No)	(1.44)	(0.35)	(0.76)	(1.66)	(12.63)
Receive restricted income	9.86***	-0.55*	-0.95	-3.68***	-35.45***
(1 = Yes, 0 = No)	(0.85)	(0.29)	(0.60)	(1.41)	(10.49)
Have endowment funds	0.93	-1.03***	-2.97***	-5.76***	-32.24**
(1 = Yes, 0 = No)	(0.78)	(0.34)	(0.68)	(1.62)	(12.71)
Income from donations	-1.339	0.128	0.356	0.001	-0.511
	(1.204)	(3.781)	(0.520)	(0.003)	(0.545)
Number of yearly reports	17.27***	4.25***	7.81***	13.32***	56.40***
	(0.53)	(0.21)	(0.49)	(0.99)	(7.39)
Number of non-zeros		-0.17***	-0.14***	-0.49***	-0.70***
		(0.01)	(0.01)	(0.03)	(0.19)
Observations	15,639	15,639	15,639	15,639	15,639
Chi-square (1)		6.66***	1.53	2.06	1.37
(p-value)		(0.01)	(0.22)	(0.15)	(0.24)

Notes. *** p<0.01, ** p<0.05, * p<0.1, robust standard errors are in brackets. Chi-square (1) statistics are for Wald test of independence (rho) of two stages. Estimates are scaled up by 1000 for ease of interpretation.

Table 2.A3. Replications using other indices

VARIABLES	χ^2			KS statistics			Deviate (1 = Deviate, 0 = Conform)		
	OLS	2SLS	Lewbel's	OLS	2SLS	Lewbel's	OLS	2SLS	Lewbel's
Charitable spending	7.412***	2.292	0.948	0.016***	-0.004	0.948	3.453***	225.851***	95.734***
	(1.245)	(13.894)	(0.612)	(0.003)	(0.036)	(0.612)	(0.560)	(65.450)	(23.685)
Interaction term	-12.473**	-90.631***	0.116	-0.037***	-0.172*	0.116	-7.240**	-562.837	-102.206
	(5.016)	(34.700)	(2.305)	(0.012)	(0.094)	(2.305)	(2.936)	(503.416)	(74.311)
Governance pressure	22.176***	86.507**	5.437**	0.062***	0.165*	5.437**	10.958***	372.345	138.010***
	(5.459)	(35.178)	(2.598)	(0.015)	(0.092)	(2.598)	(3.055)	(227.960)	(42.415)
Size	1,593***	1,299***	640***	2.218***	1.364	640***	692***	21.880***	18.348***
(logged total assets)	(160.855)	(413.861)	(97.021)	(0.390)	(1.078)	(97.021)	(100.650)	(3.677)	(2.845)
Age	153.552***	166.824***	36.973***	0.311***	0.353***	36.973***	33.305***	1.089***	1.248***
	(18.471)	(29.731)	(9.748)	(0.043)	(0.073)	(9.748)	(10.121)	(0.314)	(0.303)
Volunteers	-0.415	-0.524	-0.087	-0.002*	-0.002**	-0.087	-0.214	-0.028**	-0.035***
	(0.318)	(0.338)	(0.130)	(0.001)	(0.001)	(0.130)	(0.414)	(0.012)	(0.010)
Being audited	-2,278***	-1,924**	-280.082	-9.494***	-8.730***	-280.082	-160.785	-28.303**	-27.914**
(1 = Yes, 0 = No)	(805.031)	(849.317)	(494.184)	(2.171)	(2.263)	(494.184)	(503.643)	(14.032)	(13.753)
Receive government grants	-3,049***	-2,993***	-1,302***	-8.045***	-7.964***	-1,302***	-1,302***	-40.831***	-40.842***
(1 = Yes, 0 = No)	(515.288)	(531.932)	(312.721)	(1.269)	(1.298)	(312.721)	(322.039)	(9.549)	(9.491)
Zero fundraising	224.108	-719.428	509.013	-0.768	-3.331	509.013	494.870	22.406**	13.431
(1 = Yes, 0 = No)	(524.098)	(1,188.363)	(310.485)	(1.286)	(3.084)	(310.485)	(315.072)	(11.081)	(9.435)
Losses from investments	-232.604	-198.196	-7.282	0.924	0.961	-7.282	-164.095	0.177	3.016
(1 = Yes, 0 = No)	(759.269)	(774.008)	(405.412)	(1.657)	(1.698)	(405.412)	(420.624)	(12.644)	(12.510)
Receive restricted income	-731.273	-1,051.908*	-608.872*	-3.873***	-4.532***	-608.872*	-680.640*	-36.187***	-32.075***
(1 = Yes, 0 = No)	(581.483)	(615.156)	(337.726)	(1.411)	(1.498)	(337.726)	(348.088)	(10.556)	(10.355)
Have endowment funds	-3,010***	-3,363***	-713.194*	-5.551***	-6.561***	-713.194*	-798.128*	-26.437**	-32.996***
(1 = Yes, 0 = No)	(686.354)	(856.520)	(405.514)	(1.626)	(2.069)	(405.514)	(426.836)	(13.128)	(12.641)
Income from donations/grants	-1.339	0.128	0.356	0.001	0.006	0.356	-0.511	0.334	-0.211
	(1.204)	(3.781)	(0.520)	(0.003)	(0.009)	(0.520)	(0.545)	(0.520)	(0.545)
Number of non-zeros	-143.905***	-127.063***	-31.559***	-0.487***	-0.438***	-31.559***	-34.782***	-0.924***	-0.725***
	(10.926)	(25.305)	(5.977)	(0.025)	(0.064)	(5.977)	(6.184)	(0.223)	(0.183)
Number of yearly reports	8,093***	7,584***	1,955***	12.921***	11.383***	1,955.524***	2,076.000***	65.965***	59.227***
	(423.817)	(883.182)	(221.669)	(0.985)	(2.244)	(221.669)	(228.730)	(7.980)	(6.849)
Constant	-7.345**	0.512	78.950***	0.118***	0.143***	78.950***	77.021***	116.917	260.013***
	(2.990)	(13.383)	(1.690)	(0.007)	(0.035)	(1.690)	(1.710)	(91.868)	(49.474)

Notes: Robust standard errors for arbitrary heteroscedasticity in bracket, *** p<0.01, ** p<0.05, * p<0.1 All estimates except for χ^2 are scaled up by a factor of 1000 for the ease of interpretation. 2SLS uses the full set of our proposed instruments and treats *Interaction term* as exogenous.

2.8.7. Additional Tables

Table 2.A4 compliments Table 2.4 in the main text for estimates of control variables after various estimators. We experiment with treating *Governance Spending* and *International Term* as endogenous in specifications 2SLS (1) – (4) and Lewbel (1) – (2). We find similar results for the traditional 2SLS estimations. For the Lewbel's estimation, the endogeneity tests all fail to reject the null of statistical exogeneity of *Governance Spending.* As such, we prefer our Lewbel's estimation reported in the main analysis.

Table 2.A4. Reporting inaccuracy and NPO's observable characteristics

Variables	2SLS-1	2SLS-2	2SLS-3	2SLS-4	Lewbel-1	Lewbel-2
Charitable spending	1.676**	2.047***	2.389**	2.394**	0.156	0.167
	(0.786)	(0.704)	(1.046)	(1.201)	(0.699)	(0.854)
Interaction term	-10.761**	-9.181***	-10.693**	-10.654***	-6.018	7.592
	(4.334)	(3.155)	(4.675)	(3.595)	(5.076)	(8.022)
Governance pressure	6.143***	5.721***	6.626**	6.614***	4.382	0.604
	(2.110)	(1.872)	(2.786)	(2.441)	(3.095)	(3.875)
Observations	10,322	10,322	10,322	10,322	10,322	10,322
List of instruments						
Headcounts	Yes	Yes	-	-	-	-
Social security spending	-	-	Yes	Yes	-	-
Headcounts × Governance spending	Yes	Yes	-	-	-	-
Social Security spending × Governance Spending	-	-	Yes	Yes	-	-
Interaction term as exogenous?	No	Yes	No	Yes	Yes	No
Governance spending as exogenous?	Yes	Yes	Yes	Yes	Yes	No
K-P F-stat for weak identification	16.77	15.95	10.49	9.047	110.9	20.67
Hansen J statistics/C-statistics (p-value)	-	0.45 [0.50]	-	0.00 [0.98]	63.91/3.40 [0.00]/[0.17]	16.26 [0.00]/[0.00]
Endogeneity test for Interaction term	0.450 [0.50]	- -	4.775 [0.03]	- -	- -	2.518 [0.12]
Endogeneity test for Governance Spending	- -	- -	- -	- -	0.002 0.959	1.829 [0.18]

Note. Robust standard errors in parentheses, p-values in brackets *** p<0.01, ** p<0.05, * p<0.1. All procedures for Lewbel's estimations are performed by GMM. We report Kleibergen-Paap Wald F-statistic to account for heteroscedasticity. Overidentification test for all instruments: Hansen J statistics: The joint null hypothesis is that all of the the instruments are valid instruments, i.e., uncorrelated with the error term and that the excluded instruments are correctly excluded from the estimated equation. Tests of endogeneity: Ho: the specified endogenous can be treated as exogenous (Durbin-Wu-Hausman). Estimates are scaled up by 1000 for ease of interpretation.

Chapter 3.

Do better performing NGOs report more accurately?

Evidence from the Ugandan NGO sector*

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

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

---

* The paper benefits from discussions with Ronelle Burger, Julien Labonne, Rafaelle Miniaci, Jan Stuhler, and participants at OxDev Workshop, German Development Research Group Conference, Nordic Econometrics Meeting, Nottingham PhD Conference, and NASMES (UCDavis). This data collection was supported by the World Bank and the Japanese Government. I also thank the YITP of the IAAE, and the University of Brescia (Italy) for their hospitality during the summer 2016 when a part of this work was written.

3.1. Introduction

Non-governmental organisations (NGOs) have become an integral part of international development efforts. NGOs complement local governments by providing public goods and have become key players in channelling aid from local and international donors. In some cases, NGOs have displaced the traditional role of state institutions when donors, aiming to bypass corrupt governments, deliver aid through non-state channels (Acht et al., 2015). The sector, however, is not immune from its scandals related to misuses of funding and donations (Herzlinger, 1995; Chen, 2016). These scandals have eroded public trust and increased pressure for a higher level of financial transparency and accountability (Ebrahim, 2003). While there have been attempts to construct and validate measures to identify financial misconduct and its underlying mechanisms (see Dechow et al., 2010; Hofmann and McSwain, 2013), the literature falls short in at least two dimensions. First, there is little research on developing countries mainly due to the lack of appropriate data. Second, available measures of reporting accuracy have deficiencies, particularly when applied to small samples. To address these gaps, we use Benford's Law, a cost-effective method, to examine the level of errors in financial data of development NGOs. We study the underlying motivations of financial transparency by distinguishing between the decision to withhold data and the decision to report inaccurately when they submit all requested information. Crucially we also ask whether financial misreporting matters by investigating the connection between an NGO's performance and the accuracy of their financial data. In doing so we examine whether the widespread belief that the high cost of accounting procedures for monitoring purposes crowds out serving the client community.

To examine the prevalence of misreporting in the financial data of development NGOs we use a unique representative dataset of Ugandan NGOs (see Barr et al., 2005). Extensive information including financial data, evaluations of performance, and characteristics of the organisations was collected through interviews with NGO managers and focus groups with the beneficiary communities. 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. The tests rely on two principles. First, a dataset that contains naturally occurring numbers (without conscious manipulation) will have the first digit of all the numbers follow a logarithmic distribution shown in Hill's theorem (1995). Greater deviations between the observed data and the theoretical distribution indicate higher degrees of inaccuracy. Second, people cannot fabricate datasets that follow Benford's distribution. Psychology studies show that people instructed to invent a dataset that follows either Benford's Law or a random distribution are rarely capable of

doing so (see Schulter et al., 2010; Boland and Hutchinson, 2000). We employ several methods to measure the convergence of the two distributions which we use as the quantified degree of information accuracy. We reject the null hypothesis that the self-reported financial data conform to the Benford distribution for nearly 25 per cent of the Ugandan NGOs with complete information.

In addition to inaccuracies in the data provided, there is a second issue that the literature often overlooks: 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.[26] We ask whether non-disclosure is primarily due to unavailable financial accounts (plausibly related to a 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 is 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 is 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 Santos Silva et al. (2015). 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

[26] Woodwell and Bartczak (2008) report that 80% of surveyed donors did not provide sufficient overhead allocations to cover expenses their recipients incurred on reporting requirements.

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 be driven by the lack of resources dedicated to accounting tasks rather than manipulation to hide mediocre performance.

We further contribute to the literature on the impact of performance-based compensations on misreporting (see Burns and Kedia, 2006) by exploring the relationship between the 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 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 by using non-parametric and semi-parametric estimations. To ascertain that the mechanism is not purely caused by some unobserved motivations 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. Using non-parametric and semi-parametric estimations or an IV strategy we find the positive relationship between the degree of inaccuracy in the financial information and the NGO's perceived performance remains unchanged. That is, organisations with stronger performance are less likely to hide information and more willing to provide accurate financial data to an external body.

Our paper contributes to several branches of the literature. First, it adds a systematic study on financial reporting behaviour and performance evaluations of NGOs to the literature on the regulation and activities of watchdog/activist NGOs (Hatte and Koenig, forthcoming), and the development NGOs (Aldashev and Navarra, 2018). It contributes to the growing literature on governing and monitoring the NGO sector (Aldashev et al., 2015). Second, our measure of misreporting relates to the growing literature on forensic economics (Zitzewitz, 2012) and the use of distributional properties of numbers to identify potential misreporting in public data (see Fang and Gong, 2017 for detection of overbilling in Medicare reimbursements and Almond and Xia, 2017 for

detection of manipulation in investment returns of US non-profits). Our method also adds to successful techniques that identify and measure corruption, such as randomized auditing of officials prior to elections (Avis et al., 2016) or targeted audits by local governments (Olken, 2007). Like any measure of fraud and corruption (see Olken and Pande, 2012), digital analysis such as ours is not completely fool-proof, nor does it substitute auditing. It is, however, more cost effective and scalable to any financial and accounting accounts. Third, our exercise provides the first application in development economics to the limited number of empirical studies on distinguishing the underlying mechanism of information disclosure in labour economics (see Bettin et al., 2012 on remittances) and health economics (Dow and Norton, 2003; Madden, 2008 on cigarette consumption). Finally, our study on the link between NGOs' self-assessment and community-based evaluations adds to the literature on organisational performance and misreporting behaviour of firms (Burns and Kedia, 2006) and the recent theoretical papers on the misbehaviour and monitoring in the NGO sector (Auriol and Brilon, 2018; Aldashev et al., 2018). Different from these theoretical studies highlighting the strategic interactions within the organisations and between donors and NGOs, we propose support for a separate and novel explanation based on limited resources for irregularities in the development NGO sector, particularly regarding information disclosure behaviour.

The paper proceeds as follows. Section 3.2 describes the data. Section 3.3 describes Benford's Law and its application. Section 3.4 provides a conceptual framework to motivate our hypothesis, with a focus on distinguishing between selection models. A fuller model is provided in Appendix for Chapter 3. We discuss our econometric strategy and results in Section 3.5. Section 3.6 reports a number of robustness checks. Section 3.7 concludes.

3.2. The Ugandan NGO data

We match data from two surveys conducted in 2002 with a representative sample of Ugandan NGOs and their beneficiaries. The first survey module (NGO questionnaire) was administered to a random sample of NGOs in 14 districts of Uganda (see Map A1) 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). Ugandan field workers conducted the face-to-face interviews with an NGO representative (usually the head of the NGO).[27] Each session lasted for about two hours. The questionnaire asked 255

[27] 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

questions, including queries regarding the organisation's expenditure and income (in 2001 and 2002), funding, governance and activities the NGO undertakes. Questions regarding revenue and expenditure were designed such that the information was highly aggregated and should have been readily available from a standard annual account. If the respondent was unable to answer the questions, either the NGO representative was asked to request the information from a relevant colleague, or the enumerator left the relevant sections of the questionnaire to be completed and collected on a designated day that suited the NGO. In larger NGOs, the enumerator gave the representative the relevant section before the interview so that the NGO could prepare the required figures. We ensure that NGOs would be clearly informed about the process and be ready to answer questions on financial data. Table 3.A4 in the Appendix 3.8 provides a sample of this information. The financial data collected resembles cash flows and financial statements of for-profit firms. Ideally, we expect at maximum 60 pieces of financial information (non-zero financial transactions) to be recorded for each organisation for each year, making a total of 120 financial items over two years for our digital analysis. In the field, however, we collected much fewer data points for the analysis (43 on average). We address the implications of this fact in Section 3.4 and offer several statistical tests to accommodate the smaller sample size.

The second survey module captured community needs, characteristics and activities via a structured focus group interview conducted with beneficiary communities. In the initial NGO interview, each NGO was asked to report up to six parishes in which it had been active. One of the parishes was selected at random for a focus group interview.[28] 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.

The data collection aimed to ensure the characteristics of the focus group were not affected by the surveyed NGO by never involving the NGO in the selection of the participants or any other aspects of the evaluation. One caveat is necessary. Some of the NGOs might have favourable connections with the leaders of the parishes, who could then

---

treated confidentially." There is one group whose participants were all connected to the evaluated NGO. Excluding this group does not alter the finding.

[28] 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.

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 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.[29] 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.[30] During the data compiling process, we discovered four enumerators had exhibited "cheating incidents" while conducting the NGO interviews.[31] 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.

3.3. Using Benford's Law to detect potentially fraudulent 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). Instead of focusing on accruals and earnings, non-profit studies attempt to measure irregularities in reports by estimating either expected program 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 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 correlations between the measures and the organisation's characteristics. Third, these models require forward-looking information to construct predictions and often detailed time-series and panel data. This requirement often tempers their use in non-profit studies, in which small sample size and comprehensive data collection are the main challenges. The current literature only focuses on potential errors in some specific categories such as investment returns (Almond

[29] There is one group whose participants were all connected to the evaluated NGO. Excluding this group does not alter the finding.

[30] 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.

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

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

We use an alternative proxy for measuring the accuracy of self-reported financial data based on Benford's Law. It is a mathematical law regarding the frequency distribution of leading digits in naturally occurring datasets (e.g., the leading digit of the number 1201.17 is 1). Contrary to the basic intuition, the occurrence of each digit as the first digit is not equally likely (uniform). Instead, the first digits would follow a decreasing distribution specified in Hill's (1995) theorem. It states that random samples over different orders of magnitudes taken from a random mix of non-truncated and uncensored distributions will have the frequency of the first digits converging to the logarithmic of a distribution, dubbed the Benford distribution (See Appendix 3.8.1).[32] Financial transactions are a popular example of a random mix of non-truncated and uncensored distributions as they arise from interactions and behaviours of independent individuals. Since accurate financial records are based on repeated sums, multiplications, or quotients of prices, quantities, and financial transactions, accurately reported financial and accounting data are often expected to follow Benford's Law (Boyle, 1994).In economics, Hal Varian (1972) first suggests Benford's Law as a useful tool to uncover irregularities in accounting and financial data.[33] The premise is that those who fill in the reports and temper figures may be biased towards simpler or more natural distributions such as the uniform distribution. Psychology studies show that, whether educated or not, people instructed to make up the numbers are rarely capable of making up datasets that follow either Benford's Law or random distributions (Schulter et al., 2010; Boland and Hutchinson, 2000 for experiments involving university students). The biased report is then expected to deviate from the Benford distribution. The intensity of the bias is measured by the deviations from the theoretical distribution. Durtschi et al. (2004) outline three 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. Our NGO data satisfies all these criteria, making Benford's Law an appealing method for the type of data often available for developing country NGOs.

---

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

[33] Notable applications include Judge and Schechter (2009) to detect fieldworker cheating in household surveys; Muller (2011) to detect misreporting of Greek financial data for entry into the EU; and the detection of falsified voting in the Iranian 2009 elections (Battersby, 2009).

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. The MAD statistic is calculated by taking 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 3.A1 in the Appendix):

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

where $\mathrm{d_i} = 1, 2, \ldots, 9$ represents the first digit; $\mathrm{P_o}(\mathrm{d_i})$ is the observed frequency of digit $\mathrm{d_i}$, and $\mathrm{P_e}(\mathrm{d_i})$ is the expected frequency of digit d_i according to Table 3.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 measures 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 the degree of deviation from the Benford distribution strongly correlates with the degree of errors introduced into the financial statements.[34]

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 is:[35]

$$\mathrm{R_i} = \begin{cases} 1 - \text{MAD} & \text{if } \mathrm{C_i} = 1 \\ 0 & \text{if } \mathrm{C_i} = 0 \end{cases} \tag{2}$$

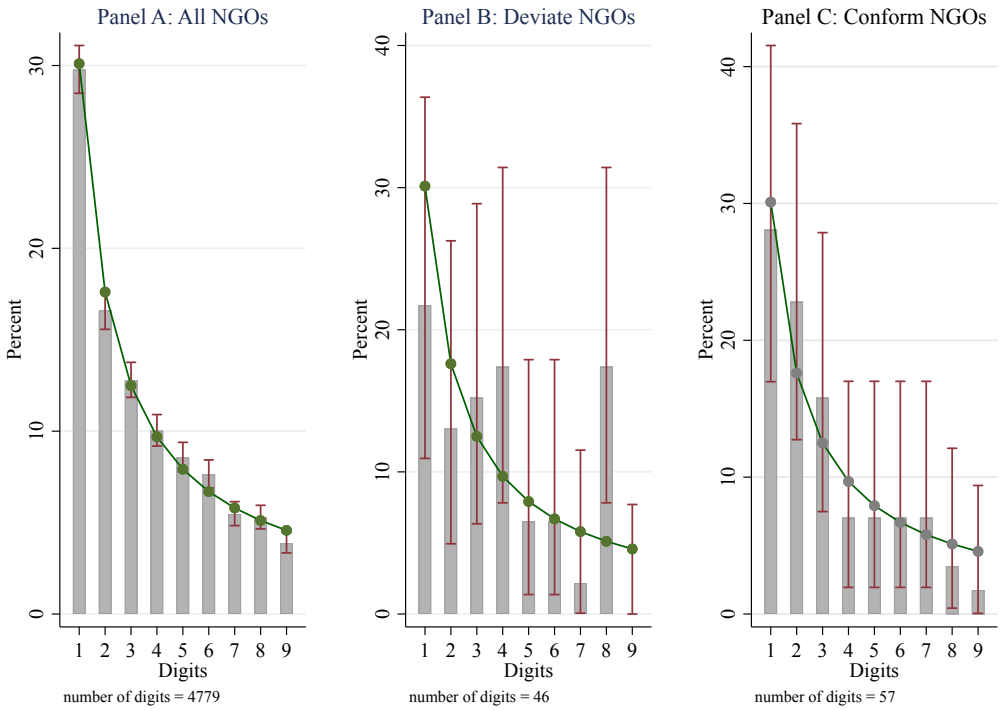
where higher $\mathrm{R_i}$ indicates that the complete report is more accurate. Since the MAD measures the level of reporting inaccuracy, we measure the reporting accuracy by subtracting MAD from the (arbitrarily chosen) number 1. $C_i = 1$ if the NGO provided all

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

[35] 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.

requested figures and $C_i = 0$ (that is, the either all the revenue figures, all the expenses figures or both, are recorded as missing variable). When the NGO reports a zero figure, we treat it as a non-missing information but zero figures are not included in the construction of MAD or our measures by Benford's Law.

Figure 3.1. The Ugandan NGO data and conformity to Benford's Law



Note. Lines represent the Benford theoretical frequency of digits 1 to 9 appearing as the first digits. Bars represent the observed distributions in the three samples. Capped spikes represent confidence intervals at the 10% significance level. In Panel A, we lump all the numbers in all financial accounts provided by the NGOs. Panel B is a representative NGO (25% of the sample) whose requested financial accounts fail the hypothesis test of conformity to the law using the two-sample Kolmogorov – Smirnov test. Panel C is a representative NGO (75% of the sample) which we fail to reject the hypothesis for its requested financial accounts.

Since there is no critical value associated with the MAD method, it could become impractical to flag potential misreporting. We employ the two-sample Kolmogorov – Smirnov to test the null hypothesis that the observed distribution of all the first digits follows the Benford distribution. We use several alternatives in Section 6. Figure 3.1 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. When an individual NGO is considered, we reject the hypothesis that the observed distribution of the first digits of all the numbers follows 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 of these NGOs.

Before proceeding, there are two concerns worth mentioning. First, we 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. Second, as the financial data were collected in face-to-face interviews, there may be concern that our result is driven by random errors introduced during the interview by either the respondent or interviewer rounding up the numbers. We believe this issue is not critical. Our measure of information accuracy relies solely on the first digit of the numbers submitted, thus alleviating the concern of the usual rounding of the last digit. To check if this is the case, we exclude from our analysis five NGOs whose reported numbers looked as if they were potentially rounded. These NGOs have more than 90% of their submitted financial numbers differing in only their first digit, followed by zeros. We obtain qualitatively unchanged results, indicating that heavy rounding does not affect our analysis. We proceed to analyse the convergence to the Benford distribution, now referred as a measure of reporting accuracy.

3.4. A Conceptual Framework

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

3.4.1. Why do NGOs provide incomplete financial information?

Despite the surveys being designed to make providing financial information easy, several NGOs provided only some of the requested figures. 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 3.1 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 discretion.[36] 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. 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). 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 the requested information 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

[36] 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.

in the first place.[37] 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 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 representative provides all *available* information during the interview regardless of the possibility of being incomplete.[38] In this case, the incomplete disclosure observed is not a strategic communication to withhold information, but a corner solution to a maximisation, subject to some *ex-ante* constraint.

3.4.2. 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 (Dye, 1985). 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

---

[37] 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.

[38] 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. To account for potential heterogeneity of the enumerator, such as NGOs responded differently to different enumerators, we include dummies of the enumerators' IDs and receive similar results.

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 have incentives to issue detailed and accurate financial records 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).[39]

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

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

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

In sum, there are two explanations for an NGO to provide incomplete information: (i) the NGO strategically withholds some information even when complete information is available; and (ii) the NGO chooses *ex-ante* to record only some necessary information. 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 consequential 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.

---

[39] Lang and Lundholm (1993) show firms with superior upcoming earnings performance have a higher propensity to reveal good news and have a higher score for disclosure quality.

3.5. Econometric Methodology and Analysis

3.5.1. Deliberate and unintended deviations from Benford's Law

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 C_i be a binary variable 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 2, 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:

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

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

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

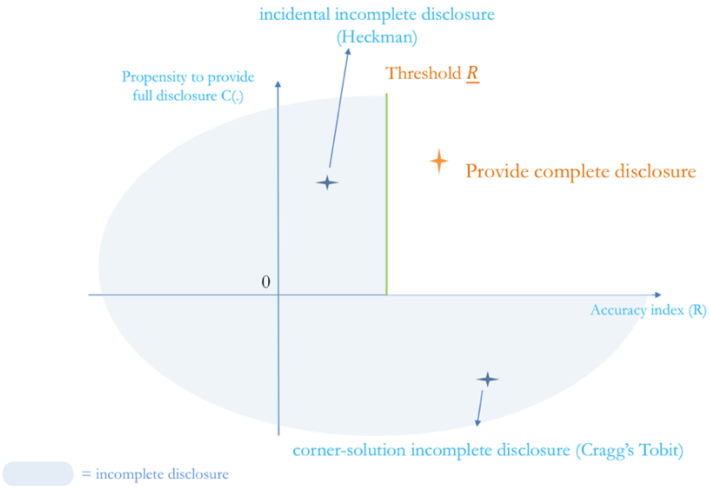
$$R_i^* = \begin{cases} R_i & \text{if } C_i > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{(Tobit model)} \quad (5)$$

In contrast, if the incomplete information is an *ex-post* strategic decision, the two stages are governed by two related mechanisms: NGOs after considering their level of reporting accuracy at 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 NGOs as latent observations, instead of genuinely zero. The appropriate model would be a Heckman selection model.

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

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 3.2 illustrates the three categories.

Figure 3.2. Three types of NGOs regarding information disclosure and report accuracy

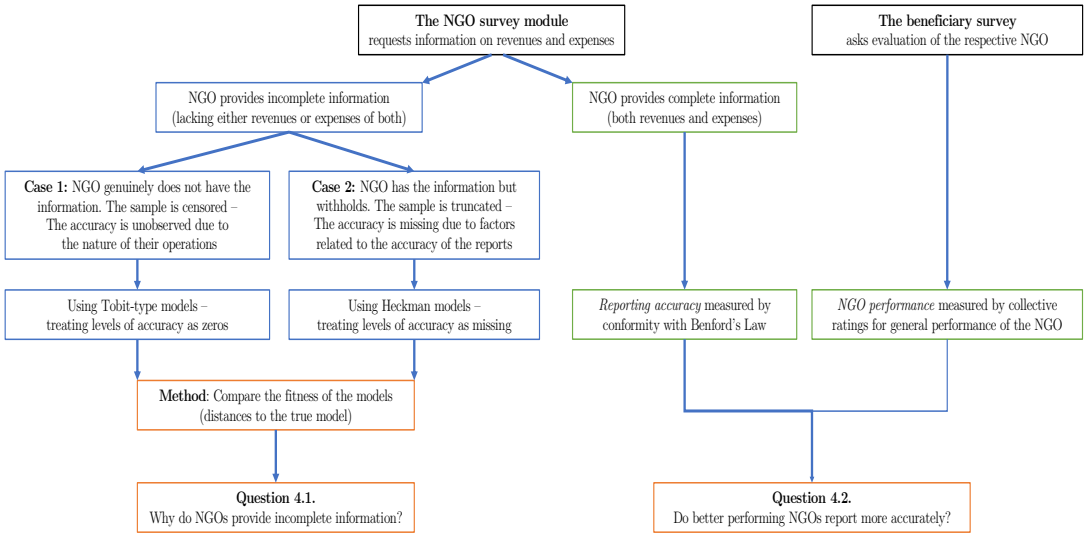


Source. Adapted from Bettin et al.'s (2012) example on remittance behaviours.

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 Santos 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.[40] Instead we use Vuong's (1989) non-nested hypothesis LR test (see Appendix 3.8.6) and a regression-based specification test by Santos et al. (2015), based on Davidson and MacKinnon (1981).[41] 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, we interpret this as a piece of evidence for strategic withholding of information. To our knowledge, this approach is the first conducted in the literature on information asymmetry.

Figure 3.3 presents an illustrative flowchart of our analysis procedure, detailing how we proceed to distinguish between the two selection models and the three types of reporting behaviour identified in our framework.

Figure 3.3. Flowchart illustrates steps of our analysis



Notes. A possible Case 3 is that incomplete information is due to the lack of authorisation or the enumerator effect. We address this concern with the survey design and checking with interviewer fixed effect in Footnote 13.

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

[41] We implement the test using command –hpc– in Stata, provided by Santos Silva et al. (2015).

3.5.2. The empirical specification

Following the hypothesis in Section 3.4, the estimation equation for the degree of accuracy and the selection model for providing all requested information are:

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

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

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).

Z_{2i} is the exclusion restriction to guarantee 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 captures for the involvement of members in the organisation's monitoring activities and is expected to satisfy the 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 randomized control trials to show 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 suggested in the literature as possible explanatory variables to the information disclosure and reporting accuracy of non-profits. To account for the concern that smaller NGOs may have fewer transactions that make their data mechanically noisier, we include *Number of non-zeros* to capture the number of non-zero financial transactions used to calculate our indices by Benford's Law.

Monitoring body: We include two variables: (i) whether the NGO has a board of directors or trustees to oversee its activities (*Oversight Board* = 1 if Yes, = 0 otherwise), (ii) whether the NGO has officially registered as a company (*Registered as Company).*

Financing source: We proxy for the accountability pressure by a binary of whether the NGO ever received a grant (*Received Grant*).

Cost of information gathering: We capture the resources needed for accounting procedures with: (i) the number of clerical staffs working for the NGOs (*Clerical staff)* and (ii) the proportion of paid employees having a tertiary education (% *Professional Degrees*).

We use: (i) *NGO Age* (the years of existence to 2002) and *NGO Age*^{2} to proxy for the NGO's reputation concern of more established organisations, and (ii) the number of years the manager has been with the NGO (*Tenure*) to examine the effect of career concerns over financial misreporting (Fudenberg and Tirole, 1995). Following Burger and Owens (2010), we use the number of reports per year required by granting bodies (*Reports per year*) to capture reporting fatigue due to heavier administrative costs.

Religious adherence: a binary variable of whether the NGO is affiliated with a religion (*Religion* = 1 if yes, = 0 otherwise). One mechanism whereby religion might influence organisational transparency is that these organisations may be frequently reminded by the (religious) moral codes of conduct (McGuire et al., 2011).

Attitude towards governance. First, we include two 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 of whether the NGO views the local government as a hindrance to their activities (*Government as hindrance* = 1 if Yes, = 0 otherwise). The variable captures 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 in the 2000s (Deininger and Mpuga, 2005) and political connection of these NGOs.

3.5.3. Descriptive statistics

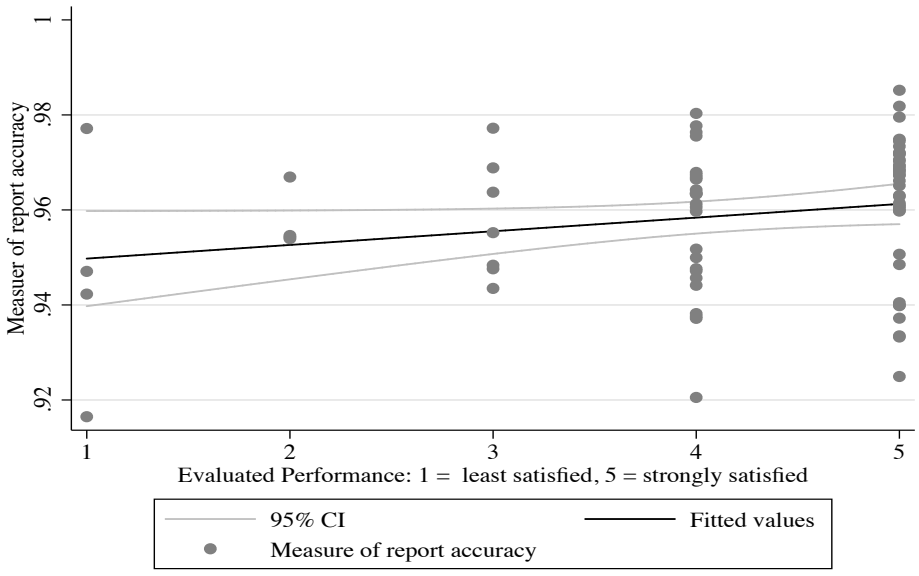
Table 3.1 presents our descriptive statistics. The number of NGOs who provided information on revenue and expenses was 77 (74% of the sample). Of these, 58 conform to the Benford distribution using the two-sample using the two-sample Kolmogorov – Smirnov test at the 10% significance level. The average 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. By way of preview, we report in Figure 3.4 the OLS estimate of interest with full control variables. The upward sloping curve suggests a positive relationship between the evaluated performance and the level of accuracy in the NGO self-reported financial figures. The estimate, however, is likely to be biased due to potential omitted variables that could explain both the report accuracy and the performance of the NGO, such as management competence and commitment.

Table 3.1. Descriptive statistics of Ugandan NGOs

VARIABLES	*Full*		*Incomplete*	*Complete*	
	Mean	S.D	Mean	Deviate	Conform
Measure of accuracy	0.71	0.423	0	0.942	0.964
Performance	4.202	1.037	4.407	3.842	4.224
Reports requested per year	1.288	2.639	0.815	2.105	1.241
% Professional degrees	0.350	0.304	0.282	0.407	0.363
Government as a hindrance	0.413	0.495	0.556	0.368	0.362
Registered as company	0.606	0.491	0.630	0.526	0.621
Received grant	0.673	0.471	0.481	0.579	0.793
Government Grant	0.154	0.362	0.037	0.211	0.189
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 connected to NGO	0.346	0.338	0.398	0.283	0.343
Observation	104		27	19	58

Note: Incomplete represents 27 NGOs who only provide either revenues or expenses related information or none. Complete represents 77 NGOs who provide all revenues related and expenses related financial figures as requested. Categorization of conformity is based on the two-sample Kolmogorov – Smirnov test between the observed distribution and the theoretical distribution.

Figure 3.4. Descriptive relation between report accuracy and performance



Note. OLS regression using the full sample from the matched 2002 Ugandan NGO survey and corresponding community data. Complete disclosure and determinants

3.5.4. Complete disclosure and determinants

Table 3.2 presents our results for the determinants of the tendency to provide an incomplete set of financial figures although the requested information was standard and should have been readily available. Having an oversight board of directors or trustees is negatively associated with the propensity to provide all requested information. One explanation is the "unintended chilling effect" suggested in Cormier et al. (2005). The oversight board may place lesser importance on a full financial report, and the NGO subject to the board agreement may have decided to keep only necessary data.[42] Being registered as a company is not significantly associated with a higher tendency to complete disclosure, raising a similar question of effective monitoring by responsible ministers in the Ugandan government office in the early 2000s (Deininger and Mpuga, 2005).

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. Regular reminders of moral codes through religious preaching is seen to be effective in promoting transparency.

We 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, managers with a longer tenure tend to have a lower propensity to full 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.

[42] 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.

Table 3.2. Estimations of the selection and outcome equations

VARIABLES	Selection equation	Outcome equation: R_i	
	C_i	Heckit	Cragg's
Performance (community satisfaction)	51.851	3.384***	3.459**
	(165.804)	(1.282)	(1.648)
Board	-849.361**	4.272	3.750
	(406.894)	(4.430)	(4.642)
Registered as company	136.415	9.316***	9.419***
	(370.485)	(2.701)	(2.675)
Received grant (Yes = 1)	808.361***	-0.740	-0.092
	(226.026)	(3.869)	(3.218)
NGO Age	139.149***	1.093	1.181***
	(51.613)	(0.681)	(0.381)
NGO Age2	-4.573***	-0.033*	-0.035***
	(1.134)	(0.020)	(0.010)
Tenure length	-90.388*	-0.075	-0.112
	(51.517)	(0.336)	(0.186)
% Professional degrees	1,287.665**	-1.015	0.197
	(598.840)	(6.155)	(2.903)
Clerical staff (log)	75.902*	0.019	0.036
	(41.018)	(0.205)	(0.143)
Reports requested per year	48.942	-0.923*	-0.846**
	(78.927)	(0.557)	(0.404)
Religious Affiliation	1,574.535***	-4.757	-4.062
	(334.990)	(3.818)	(2.540)
Lack of skilled staff	231.249	-1.997	-1.732
	(336.440)	(2.794)	(2.644)
Lack of funding	-503.696	2.564	2.351
	(369.878)	(2.937)	(2.385)
Government as a hindrance	-739.811**	-2.261	-2.760
	(298.481)	(3.249)	(3.329)
Years working in government	2.964	-0.462**	-0.444*
	(16.210)	(0.174)	(0.231)
Member involvement	762.451**		
	(367.103)		
Number of non-zeros		0.543***	0.549***
		(0.103)	(0.076)
Lambda/sigma		-2.522	10.357***
		(8.924)	(0.647)
Constant	-221.743	912.200***	909.986***
	(733.631)	(10.870)	(9.545)
Observations	104	104	104

Note. *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (14) in parentheses. Estimates are multiplied by 1000 for easier interpretation. Jaque-Bera test statistics of residual normality in the outcome equation: Pr(Skewness) = 0.142, Pr(Kurtosis) = 0.3699, Joint-test chi-square statistic (p-value) = 3.03 (0.22). Breusch-Pagan heteroscedasticity statistics (p-value) = 3.91 (0.05). Variable Performance is the evaluation from the community focus group. Section 3.5.2 provides a description of the controls.

We find 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. 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 to keep all the standard financial data. NGOs whose managers consider government as a hindrance to daily operations are less likely to provide all requested 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 a 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 3.A6 in Appendix 3.8.6, 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 measures for resource constraints may already have absorbed any significant effect of more substantial information gathering costs due to frequent donor requests. Distinguishing the censoring mechanisms is necessary to explain the insignificance.

Table 3.3 provides Vuong's (1989) and the HPC (Santos 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 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 that the NGO would withhold the financial figures if their data accuracy fails to exceed. Section 3.5.4 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.

Table 3.3. 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 (H1)	Cragg's specification is closer			
Ln Ration (s.e) [p-value]	6.753 (0.394) [0.000]			
Observations	104	104	104	104

Note. Both tests unanimously indicate the double-hurdle censoring mechanism fits the data better. See Appendix 3.8.6 for details of the tests.

Although we statistically support that a corner solution censoring mechanism fits the Ugandan data better, the two models 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 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.[43]

[43] 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 as 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.

3.5.5. 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. The decision to disclose all requested information is not related to the performance of the organization but rather their capacitary and human constraints. The decision to report accurately, however, relates to higher ratings of performance, perhaps to reveal their type or avoid punishment of being detected as untruthful as in Benabou and Tirole (2006).

The 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 age 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. Another supporting evidence is that a higher number of reported financial transactions (*Number of non-zeros*) also positively correlates with the report being more likely to conform with the Benford distribution.

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 the 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 a 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 3.1 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 3.6.3), the negative association remains (see Table 3.A3).

3.6. Robustness checks

We show that our results are not driven by alternative measures of reporting accuracy, potential endogeneity between performance and reporting accuracy, and the functional form and distribution assumptions of the specifications.

3.6.1. 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 3.8.2 presents three "critical-value" based measures widely used in previous studies (for example Michalski and Stolz, 2013; 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 Barabesi et al. 2018 for discussions). Table 3.A2 reports the results for each of the alternative test statistics.

3.6.2. Endogeneity between performance and reporting accuracy

The main analysis accounts for sample selection by a Heckman correction model. Omitted variable bias could be another concern. 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 sample selection and censored Tobit estimators with endogenous regressors proposed by Smith and Blundell (1994), Wooldridge (2010), and Semykina and Wooldridge (2010) (see Appendix 3.8.6 for details).[44]

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 one, NGO characteristics are unlikely to satisfy the exclusion restrictions as they could be correlated with the unobserved NGO ability. For another, 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 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. staff or an NGO member (%).[45] Our identification assumption is that certain demographics of the focus group (connection with the NGO and the average age of focus group members) could strongly predict the (evaluated) performance of an NGO. Groups with more respondents with a connection to an NGO may report higher satisfaction because the NGO has exerted effort in addressing specific needs of the community. Similarly, if the needs of a particular demographic are less well-served, we would expect a focus group with a higher proportion of this demographic to result in lower satisfaction scores. The intuition comes from documented institutional characteristics of the NGO sector. Barr and Fafchamps (2006)

[44] In Table 3.2, 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. Some NGOs could have the reservation level so low that they choose 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 omitted heterogeneity in Table 3.3, the correlation becomes positive (Lambda = 3.217) as predicted.

[45] 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.

and Fruttero and Gauri (2005) show that many NGOs are often less inclined to work in remote and poorer communities despite these places often being the neediest and requiring the most attention. Evidence suggests some NGOs do not always try their best to serve the community, resulting in heterogeneous degrees of communities' satisfaction. Likewise, specific to Uganda there is evidence that no NGO reported any service or activity that specifically targeted the elderly (Barr et al., 2003). Using this intuition, we expect that groups with more respondents with a connection to an NGO would report higher satisfaction because the NGO has exerted effort in interacting with their community. Also, we expect NGOs working in areas with a larger proportion of elderly would receive a significantly lower satisfaction score since the needs of this group are less well-served. F-tests and coefficients in Panel B in Table 3.3.4 confirm our conjectures. Groups with more participants having a connection with the evaluated NGO tend to give higher scores; whereas groups with more senior-age members give unfavourable assessments – significant at the 5 and 1 per cent level, respectively.

We believe the instruments are exogenous to the accuracy of the financial report for three reasons. First, there is no reason why a demographics of the community, like age, could affect the tendency of an NGO to report accurately to a third party. The demographics could affect how transparent the NGO is towards the village, but hardly with a third party (here the enumerator). Second, as explained in Section 2, the focus group selection can be considered exogenous to the concerned NGO. The community leader selected the participants without knowing the NGO that was to be evaluated. It was therefore not possible for the leader to choose a focus group that would give a biased evaluation of a specific NGO. Indeed, Table 3.A3 shows the instruments do not exhibit any significant correlations with observable characteristics of the NGO working in the community. The balanced tests also 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 be able to self-select into villages by using the demographic statistics used as our instruments. Unlike other community characteristics relating to infrastructure and prosperity of the communities 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 3.4 provides Hansen J statistics and Anderson-Rubin Wald test for the null hypothesis that the orthogonality conditions are valid. In both tests, we fail to reject the null hypothesis that our instruments can be excluded from the main equation.

Panel A in Table 3.4 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 coverage-corrected confidence intervals in Table 3.4 (see Cameron and Miller, 2015 for recommendations). The confidence interval is [0.61, 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 measurement error, implying that NGOs with stronger performance in serving the communities would tend to provide financial reports more accurately.

Table 3.4. IV estimations of the selection and outcome equations

Panel A. Estimates of the selection and outcome questions

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

Panel B. Diagnostic test for IV first stage estimation

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

Note. *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (15) in parentheses unless stated otherwise (bootstrapped standard errors with 100 replications are largely identical). The coefficient θ_2 of the predicted value of the suspected endogenous variable ($\hat{v}$) = -9.212 * (5.160), indicating the presence of endogeneity.

3.6.3. 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. Figure 3.A1 in Appendix 3.8. reports a clear graphical resemblance between the predicted residual density and the normal distribution. In Table

3.2, the Jaque-Bera test statistics strongly supports residual normality, while we barely fail to reject the null hypothesis of homoscedasticity at 5%. In this section, we relax the assumptions on functional forms and distribution types of the error terms, showing our result for the positive relationship remains.[46]

We estimate two simple nonparametric kernel regressions (local polynomial smooth and lowess smother) in Figure 3.A2 for a bivariate relationship between performance and reporting accuracy.[47] The nonparametric estimation becomes complex when adding more variables into a kernel regression as it introduces locally sparse noises ("curse of dimensionality").[48] 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 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 3.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 Appendix 3.8.6 for details). The estimators are robust to heteroscedasticity, consistent, and asymptotically normal for a wide class of error distributions. Figure 3.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 the district level, and bias-corrected for $Performance_i$ are 5.381, 3.318 and [4.616, 16.517] respectively. The results

---

[46] In the main analysis, we estimate the Cragg's (1971) normal truncated model, which assumes a normal distribution of the report accuracy variable. It is unsatisfactory to have the possibility that a rational NGO might plan to have a negative degree of report accuracy. 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. As such, we report results from the normal truncated model.

[47] 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.

[48] 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.

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.

3.7. Concluding remarks

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

We also find that the underlying mechanism for nondisclosure of financial accounts and the provision of false information is 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 mobilize the donor community to contribute towards funding such oversight mechanisms. This mobilization could help harmonize the accountability pressure on organisations with limited resources, especially when under-staffed NGOs are unable to accommodate both community services and expensive bookkeeping procedures. As it becomes more difficult for NGOs that fail to provide credible financial reports to receive funding in the future, the limited resources devoted to accountability tasks could lead to a vicious cycle within the NGO field. Over time, oligopolisation of the NGO sector could become a serious problem as donors could move away from under-staffed NGOs due to concerns over financial transparency. Designing specific financial packages for overhead costs could be the first step to improve the accountability of the sector and maintain the benefits that come with grassroots NGOs.

We also suggest that higher community satisfaction scores are aligned with accurate reporting, challenging the widely held belief that upward and downward accountability could conflict. Given the alignment 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. We also find a strong correlation between excessive reporting requirements and lower levels of accuracy. The implication is that when a reasonable reporting burden is exceeded, cynicism may erode the commitment to accurate and transparent reporting of organisations with limited resources.

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.

3.8. Appendix For Chapter 3

3.8.1. The Benford's distribution of first digits

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

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

where P(d) is the probability that digits $d = 1, 2, \ldots, 9$ occurs as the leading digit in a naturally drawn set of numbers. Table 3.A1 records the full theoretical distribution specified by Benford's Law.

Table 3.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

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

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

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

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

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

where N is the total number of non-zero financial items used, $P_o(d_i)$ and $P_e(d_i)$ are respectively the frequency of digit d_i appearing in the observed sample and the expected (theoretical) distribution, $c(\alpha) = \sqrt{-\frac{1}{2}\ln(\frac{\alpha}{2})}$ is the critical value of the two-sample Kolmogorov – Smirnov test at N and test power $\alpha = 0.10$ (for which $c(\alpha) = 1.22$). For the Cho-Gaines' (2007) *d*-statistics, we use 5 instead of 1 to construct the measures based on Cho-Gaines' *d*-statistics since the statistics are larger than 1 (see Morrow, 2014 for the critical values). The number 5 is arbitrarily chosen to facilitate the computation of the Cragg model and varying its magnitude does not change the results. We also transform the usual KS statistics (that measures the divergence of the two samples) into the formula above to measure the convergence of the two samples. Like the measure from the MAD statistic, lower values of the indices indicate that the tested data diverge further from the Benford distribution. Table 3.A2 show similar results for the main equation, implying that our results are not sensitive to the construction of the indices.

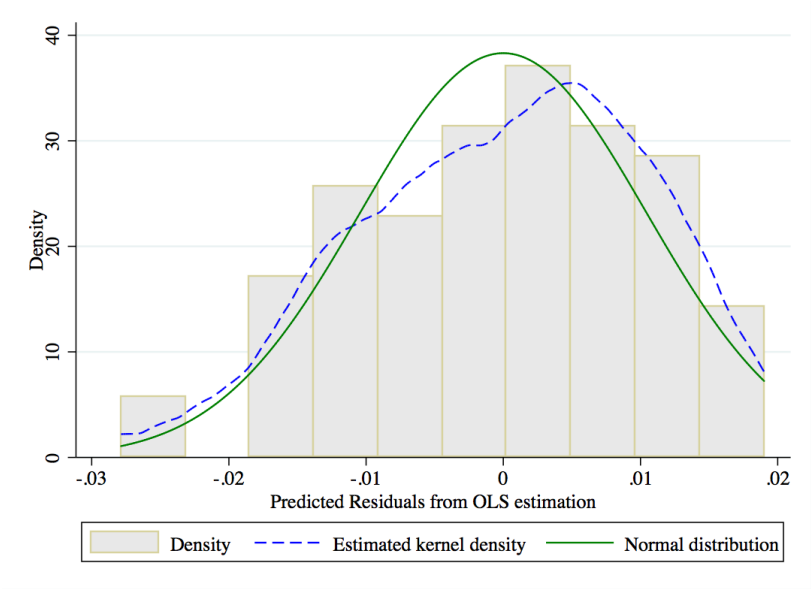
Table 3.A2. Estimations of outcome equations for alternative measures of accuracy

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

Note. *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level in parentheses. Estimates are multiplied by 1000 for ease of interpretations.

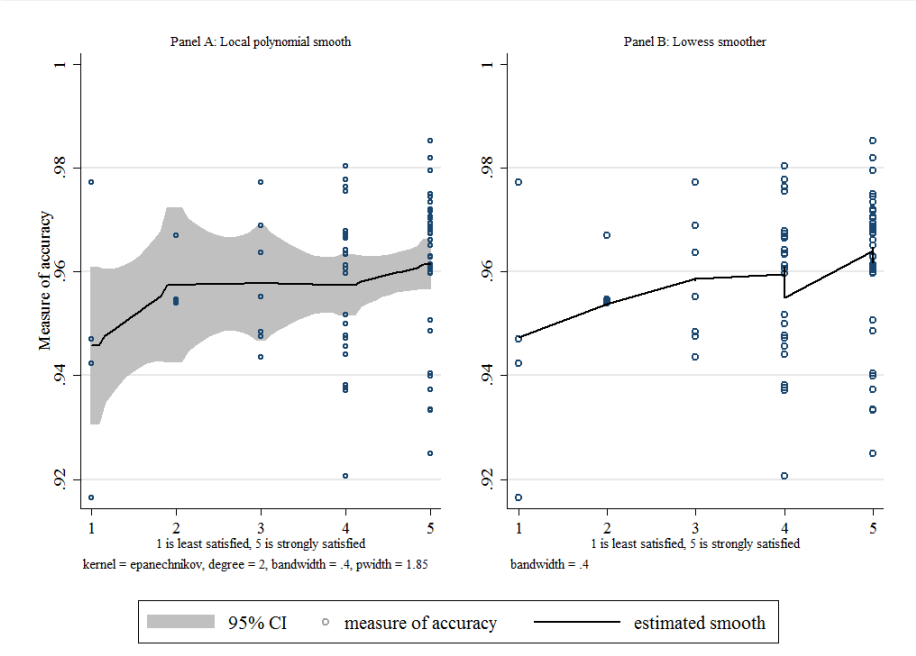
3.8.3. Non-parametric and semi-parametric estimations

Figure 3.A1. Density plot of predicted residuals from OLS of the main equation



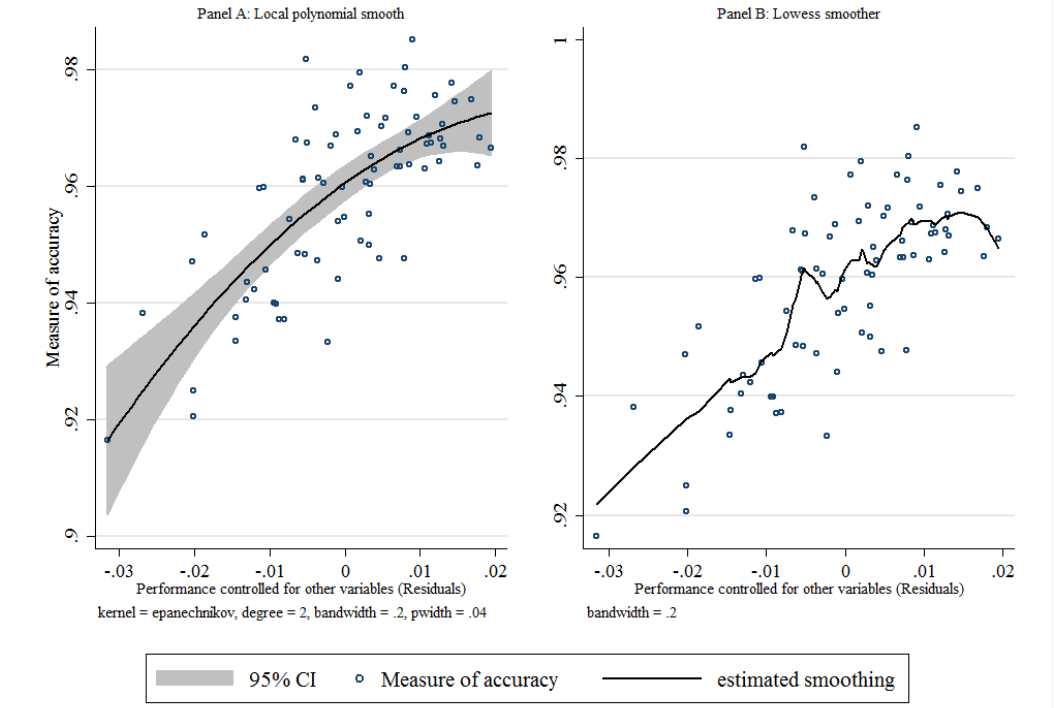
Note. Estimated kernel density shows a clear resemblance with the normal distribution, implying the normality condition in residual terms is graphically valid.

Figure 3.A2. Kernel-weighted regressions of accuracy (y) and NGO performance (x)



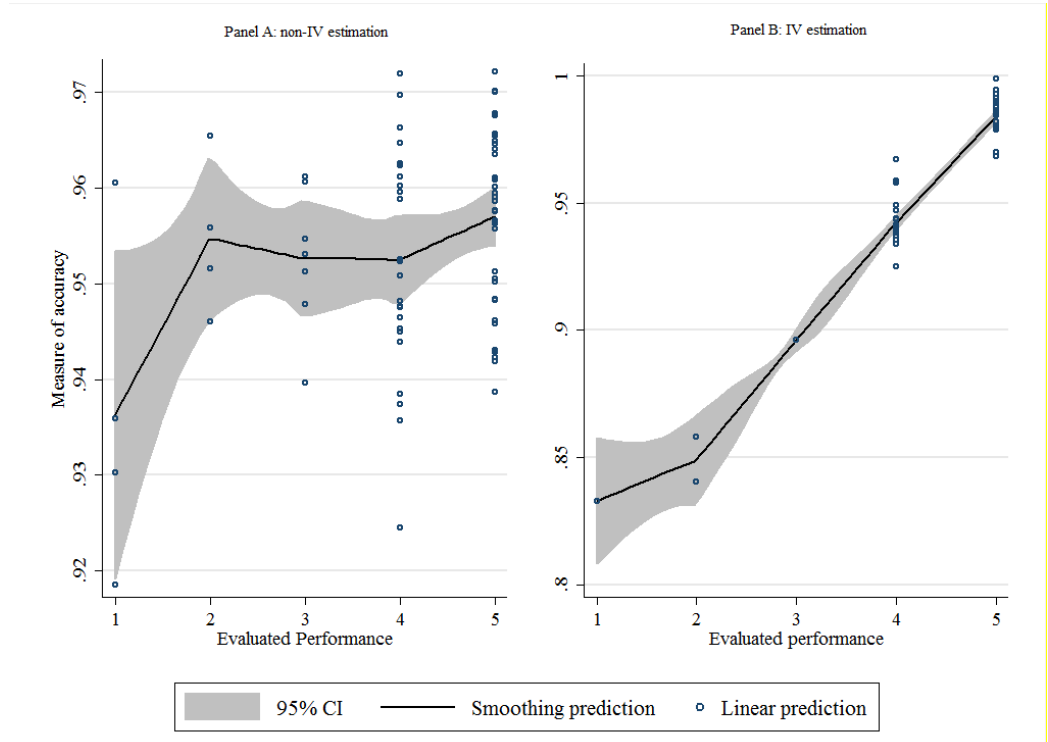
Note. Panel A reports a kernel-weighted local polynomial regression (multiple piecewise linear estimations *locally* in a neighborhood of x within a given bandwidth). Polynomials of x are included to improve the fit of the estimation. The default Kernel (Epanechikov) distribution and 95% confidence interval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

Figure 3.A3. Kernel-weighted regressions of accuracy and performance (residuals)



Note. 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 3.A4. Semiparametric non-IV and IV estimation following Robison (1988)



Note. See Appendix 3.8.4 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.

3.8.4. Balance tests for excludability argument of the instruments

Table 3.A3. A balance test for NGO characteristic and instruments

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

Note. *** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses. 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.

3.8.5. A Sample of Financial Information Requested in the 2002 survey

Table 3.A4. Sample of Financial Information Requested in the 2002 survey.

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

Grants/voluntary donation/contribution to another NGO/CBO/Church		,000	,000
Interest on debt and financial charges (including leasing)		,000	,000
Bribes		,000	,000
Other costs		,000	,000
B. Investments		,000	,000
Land and buildings		,000	,000
Vehicles		,000	,000
Equipment and machinery (including computers)		,000	,000
Total annual expenditures		**,000**	**,000**
BORROWING SITUATION			
Loans/borrowed money from:		,000	,000
Ugandan/International NGO	Borrowed	,000	,000
	Remaining	,000	,000
Local/national Government	Borrowed	,000	,000
	Remaining	,000	,000
UN agency/Bilateral donor	Borrowed	,000	,000
	Remaining	,000	,000
Bank/financial institution	Borrowed	,000	,000
	Remaining	,000	,000
OTHER FINANCIAL INFORMATION			
Amount of Investment financed by hire-purchase		,000	,000
Values of equipment		,000	,000
Values of inventories		,000	,000
Values of vehicles		,000	,000

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

3.8.6. Technical details

3.8.6.1. A model 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 maximization problem:[49]

$$\max_{P_i,R_i} U_i(P_i,R_i) \;\; s.t \; E_i = P_i + R_i + \tau C(.) \tag{A6}$$

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.[50] 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 report 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

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

[50] 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.

the NGO spends $R_i = 0$, we have $C(.) = 0$ or *incomplete disclosure.* Reporting with a degree of R_i 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 $R_i > 0$.[51] The parameter $\tau > 0$ represents the fixed cost of disclosing full information, which includes either costs of information acquisition or ex post costs discussed above.

To reflect the altruistic aspect, we further assume:

- $U_i(0, R_i) = 0$ and $U_i(P_i, 0) > 0$ and $U_i(P_i, .)$ is a strictly quasi-concave function of P_i. This set ensures that the altruistic NGO derives no utility from diverting all the given resources away from delivering core projects and they always gain from completely focusing on altruistic activities. The last assumption is to imply the NGO's preference on the consumption set of performance measures (such as beneficiary's feedback) is convex.[52]
- $U_P^i = \frac{\partial U_i}{\partial P_i} > 0$, $U_{PP}^i > 0$, $U_R^i = \frac{\partial U_i}{\partial R_i} \geq 0$, $U_{RR}^i > 0$. 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).[53] 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 $R_i^* \leq \underline{R}$.

---

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

[52] 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.

[53] 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.

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

Proof. If $\mathrm{U_R^i} = 0$, the utility function is constructed such that the NGO's preference does not attach any additional utility to either increased accuracy or complete disclosure. The problem has a unique solution $(E_i, 0)$.[54] In this case, the construction of U_R^i governs both mechanisms underlying disclosure and accuracy decisions.

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

$$\max_{P_i} U_i(P_i, E_i - \tau - P_i) \quad s.t \qquad E_i - \tau \geq P_i \geq 0 \tag{A7}$$

$$\max_{P_i} U_i(P_i, E_i - P_i) \qquad s.t \qquad E_i \geq P_i \geq 0 \tag{A8}$$

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

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

$$G(tx_1 + (1-t)x_1') = tG(x_1) + (1-t)G(x_1') \geq 0 \tag{A9}$$

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

$$U_i(tx_1 + (1-t)x_1', .) > \min\{U_i(x_1, .), U_i(x_2, .)\} = u_1 \tag{A10}$$

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

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

First, if $\mathrm{u_1} = \max_{E_i - \tau > P_i \geq 0} U_i(P_i, R_i) < U_i(E_i, 0)$, 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

[54] Another situation 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 $U_P > 0$.

$\lim_{R_i \to 0} U_i(P_i, R_i) < U_i(E_i, 0)$. 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 maximizes utility by choosing $\mathrm{R_i} = 0$. *QED*

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

Proposition 2. If $\mathrm{R_i^*} \in \mathbb{R}_{\geq 0}$ is the optimal 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 accuracy decreases.

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

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

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

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

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

$$KLIC \equiv E(L^{True}) - E(L^*) \tag{A11}$$

where L^{True} is the log of the conditional density of the unknown true model and L^* is the log of the conditional density of the model approximating the data. Vuong (1989)

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

$$z = \frac{LR_n\left(\hat{\beta}_n, \hat{\theta}_n\right)}{\hat{\omega}_n\sqrt{N}} \equiv \frac{\left[L_n^{Cragg}\left(\hat{\beta}_n\right) - L_n^{Heckman}\left(\hat{\theta}_n\right)\right]}{\hat{\omega}_n\sqrt{N}} \tag{A12}$$

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:

$$\hat{\omega}_n^2 \equiv \frac{1}{n}\sum_1^n \left[\ln\frac{f\left(\hat{\beta}_n\right)}{f\left(\hat{\theta}_n\right)}\right]^2 - \left[\frac{1}{n}\sum_1^n \ln\frac{f\left(\hat{\beta}_n\right)}{f\left(\hat{\theta}_n\right)}\right]^2 \tag{A13}$$

where $f\left(\hat{\beta}_n\right)$ and $f\left(\hat{\theta}_n\right)$ 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.

3.8.6.3. Cragg's and Heckman model with endogenous regressors

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

$$R_i^* = \alpha_1 + \gamma x + X_i'\beta_1 + u_i \tag{A14}$$

$$C_i(.) = \alpha_2 + + S_i'\beta_2 + v_i \tag{A15}$$

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 maximum likelihood function specified in Cragg (1971). Another advantage of the Smith-Blundell procedure is it does not require any distributional assumptions for the first stage estimation. For these results, we adopt Blundell-Smith procedure by MLE as our IV-Cragg estimator (Wooldridge, 2010 p. 682): first, estimate the reduced form of x on instruments Z_i and other exogenous variables by OLS: $x = Z_i'\gamma_2 + S_i'\alpha_2 + u_{2i}$. Obtain estimated parameters of the OLS residuals of x as: $\hat{v} = x - Z_i'\hat{\gamma}_2 + S_i'\hat{\alpha}_2$. Second, estimate a standard Cragg's double-hurdle model with the main equation as: $R_i^* = X_{1i}'\beta_3 + \beta_{\text{IV-Cragg}}x + \theta_2\hat{v} + v_{2i}$. The standard errors are clustered at district level and bootstrapped with 100 replications.

The estimates $\beta_{\text{IV-Heckman}}$ and $\beta_{\text{IV-Cragg}}$ 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; and (v) Hansen J statistic test for overidentification. We note these statistics are not technically equivalent for nonlinear models and only provide informative indications (see Sanderson and Windmeijer, 2016).

3.8.6.4. Robinson (1988) and Powell (1984) estimators

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

$$R_i = m(x_i) + X_i'\beta + \lambda(S_i) + \varepsilon_i \tag{A16}$$

where 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:

$$\underbrace{R_i - E[R_i|x_i]}_{u_1} = (X_i - E[X_i|x_i])'\beta + \lambda(S_i) - E[\lambda(S_i)|x_i] + \varepsilon_i = \underbrace{(T_i - E[T|x_i])\gamma}_{u_2} + \varepsilon_i \tag{A17}$$

To avoid imposing any functional form, the estimator replaces the unknown quantities $E[R_i|x_i]$, $E[X_i|S_i]$ 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 address the case that $\hat{\lambda}(S_i)$ does not converge to its parametric estimation.[55] We also experiment with different trimming levels incrementally from 0.00 to 0.05. Since the results are 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 $\varepsilon_i = \rho\tau + \eta_i$. The selection model becomes:

$$R_i = m(x_i) + X_i'\beta + \lambda(S_i) + \rho\tau_i + \eta_i \tag{A18}$$

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

$$\underbrace{R_i - E[R_i|x_i]}_{u_1} = (X_i - E[X_i|x_i])'\beta + \lambda(S_i) - E[\lambda(S_i)|x_i] + +\rho(v - E(v|x_i)$$

$$= \underbrace{(T_i - E[T|x_i])\gamma}_{u_2} + \varepsilon_i \tag{A19}$$

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

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

$$\hat{\alpha}_{CLAD} = \text{argmin } Q_n(\alpha) = \text{argmin } \frac{1}{n}\sum_{i=1}^{n}[R_i - \max\{0, X_i'\alpha + \varepsilon\}] \tag{A20}$$

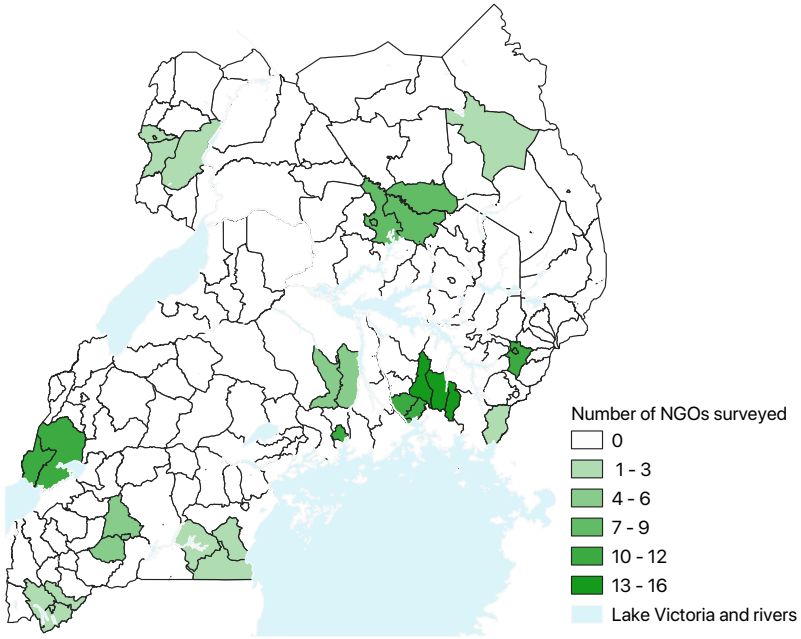
Bunchinsky (1998) provides an iterative linear programming algorithm (IPLA) to computationally estimate the parameters.

---

[55] The clustered variance is $V(\hat{\gamma}) = \left(\hat{u}_2'\hat{u}_2\right)^{-1}\sum_{j=1}^{n_c}\theta_j\theta_j'\left(\hat{u}_2'\hat{u}_2\right)^{-1}$, where $\theta_j = \sum_i \hat{u}_i t_i$, $\hat{u}_i$ is the residual for the ith observation and t_i is the row vector of T_i; $n_c = 14$ is the number of clusters.

3.8.6.5. The geography of the surveyed NGOs in 2002

Map 3.A1. Districts in the 2002 Ugandan NGO survey.



Note. Darker colours indicate more NGOs were drawn from the districts.

3.8.6.6. Robustness to different ways to measure performance of NGOs

In the main analysis, the focus groups were asked to jointly give a rating about the satisfaction of the people who live in the community (parish) with the concerned NGO on a Likert scale (1 = least satisfied and 5 = most satisfied). As the rating was collectively decided among the participants and the participants were selected independently of the NGO, we aimed to minimise the subjective bias by each individual respondent. There, however, may still exist measurement errors in the dependent variables. It is not a serious concern in our study. Indeed, we are not concerned that the measurement error being correlated with the outcome of interest (the non-classical measurement error) for two reasons. First, there were several NGOs working in each parish (community) and the respondents and the community leader were not told which NGO to be evaluated. Second, the outcome of interest (the accuracy of the reported financial data) is calculated by a statistical procedure and the focus group participants knew beforehand about neither the financial data requested in our NGO interviews nor the fact that the concerned NGO was interviewed. It is, therefore, difficult to believe that the participants were selected in a way that correlates with the way that the NGO reported their financial data. As such, we are left with the concern of classical measurement error, which results in attenuation bias, that is our estimates of interest are closer to zero than the true, unobserved values. Given our estimates of interest are significant and positive, this is not of concern for our inference.Nevertheless, the focus group questionnaire did have several other relevant

aspects such that how good the NGO is with what they do (*NGO_good*); how important the NGO is with the community's lives (*NGO_important*), how accessible the NGO is to the community (*NGO_assessible*), and how quickly the NGO responds when the community asks for help (*NGO_quick_respond*). For completeness, we report in Table 3.A5 the correlations of our preferred measure and the four measures as well as in Table 3.A6 checks using these measures as the alternative independent variables. We note that the alternative measures are positively correlated with our preferred measure and the estimates are positive, despite less precise. These exercises suggest that using other aspects of satisfaction of the general performance does not alter out main results.

Table 3.A5. Correlations of closely relevant measures

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

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

Table 3.A6. Robustness to different measures of Performance

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

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

Table 3.A7. Estimates of selection and outcome equations for control variables

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

Note. *** p<0.01, ** p<0.05, * p<0.1 Clustered standard errors at district level (14) in parentheses (bootstrapped standard errors with 100 replications are largely identical). Table 3.F3 presents estimates of control variables for the IV estimation. Estimates are multiplied by 1000 for easier interpretation.

Chapter 4.

What motivates development NGOs?

Evidence from the Ugandan NGO sector*

Keywords: NGOs, diversification, risk aversion, motivations, aid, flood
JEL Classifications: L25, L31

* This chapter benefits from discussions with Gani Aldeshev, Torben Fischer, Andreas Ruiz-Larrea, Hector Rufancos, Daniel Siedmann, Alex Possajennikov, and participants at CREDIT seminar (Nottingham), Sussex, the 2016 RES Symposium (Brighton), and the 3rd IAAE Annual Conference (Milan), German Research Group in Development Economics (Zurich), SSES Annual Meeting (Gothenburg). I especially thank Ronelle Burger; without her support and effort in collecting the 2008 dataset, this paper would have not been possible. This paper is loosely based on and supersedes the preliminary work submitted for my MSc in Economics and Development Economics I undertook at the Nottingham School of Economics. The current paper expands the theoretical and empirical framework to significantly differentiate from the previous study.

4.1. Introduction

Development NGOs have become an essential part in channelling aid and delivering development projects. There has also been mounting interest from the public, governments and donors in the motivations underlying the work that NGOs do. Understanding such motivations help guide the interested parties on how best to monitor and support the sector to ensure aid is effectively delivered. The existing literature on organisational incentives and prosocial behaviour yet remains inconclusive as to what determines the decisions made by NGOs (Watkins et al., 2012). One particular area of interest is the decision of an NGO to specialise in one or to diversify into other activities (Kistruck et al., 2013). For one, the shift towards engagement in multiple activities could accommodate a broader range of social demands and beneficiaries (Edwards and Hulme, 1992, p.80). However, pursuing activity diversification is costly and potentially harmful to the development effort due to the lack of focus which results in wasted transaction costs and management inefficiency (Tirole, 1988; Farrington and Bebeington, 1993). More importantly, involvement in multiple roles and directions could be interpreted by the public as "mission vagueness" that could diminish the reputation and legitimacy associated with the NGO status. In this paper, we examine underlying the motives that shape an NGO's decision to diversify their activities. In particular, we focus on two camps often raised by academics and development practitioners, which we refer to as to reduce idiosyncratic risk to funding and to derive personal gains.

The risk reduction camp focuses on the mitigation effect of diversification against the idiosyncratic uncertainty that is inherent in NGOs' funding streams. First, if NGO's funding is granted according to an evaluation of their performance in a particular task, they may spread their effort to multiple activities to reduce the uncertainty surrounding such measurements (Montgomery, 1994).[57] Second, Kistruck et al. (2013) and Keating et al. (2005) provide descriptive evidence of how NGOs in developed countries use income diversification both as a safety net and to improve financial stability. In addition to uncertainty reduction, NGOs might diversify to enhance their own personal gain, namely, careerism (Gibbons and Murphy, 1992), rent-seeking (Murphy, 1999), "managerial entrenchment" (Shleifer and Vishny, 1989, p.137), egoistic gains and prestige (Olson, 2009; Benabou and Tirole, 2006), and impure altruism (Andreoni, 1998). We further discuss these motivations in Section 4.2.

[57] An observable measure of the organisation's performance. In our context, it could be the number of people that an NGO was able to help during a humanitarian campaign.

To understand how these motivations interact in determining an NGO's decision, we incorporate them into a multitask principal-agent model with an incentive contract between a stakeholder and an NGO. We show that the value-based incentives given by the donor through contracts (such as grants from donors, the membership fees or user fees paid by beneficiaries) and the NGO's optimal level of activity diversification would depend on five parameters: the variance of idiosyncratic uncertainty associated with the NGO's performance, the NGO's risk aversion, the disutility of eliciting effort, the marginal cost of diversification, and the magnitude of the private gains from diversification. Our comparative statics at equilibrium allow for two scenarios. When the magnitudes of parameters related to the motivation to reduce risk (the variance of the uncertainty and the risk aversion) dominate those of the other parameters, the optimal diversification level decreases with the value-created incentives set by the stakeholder. The intuition is that when risk reduction is the key motivator in the NGO's decision-making process, further value-created incentives would lead the NGO to focus on value-created task instead of diffusing effort into a more complex portfolio. In contrast, when the magnitudes of parameters related to private gains dominate, the NGO would diversify more following an increase in the incentive. It is because any extra incentive from the stakeholder now works as insurance to guard against idiosyncratic risks and uncertainty surrounding their future funding, allowing the NGO to diversify more to gain personally.

These results provide important empirical implications. If we find NGOs respond to an increase in incentives from stakeholders by lowering their diversification level, our model suggests that the negative effect of incentives on diversification is due to NGOs having a higher need for risk reduction (for example, a higher risk aversion or facing a higher risk variance). In contrast, a positive effect of incentives on diversification implies that deriving personal gains might drive the NGOs' decision regarding their activities.

We examine the above predictions using a unique representative dataset of 402 Ugandan NGOs collected in 2008 by Burger and Owens (2013). We measure activity diversification using two distinct methods: the number of activity categories (our preferred measure) and a Herfindahl-Hirschman index of spending on the activity categories at the end of 2007. To capture the contractual relationship between the NGO and its stakeholders, resembling the setup in the theory, we use the reliance of an NGO's income on contractual sources: the ratio of revenues from stakeholders, namely, international and local grants, membership and user fees, over the total revenue in 2007. We exclude voluntary donations and business incomes from the measure as NGOs rarely have contractual responsibility attached to these sources of finance.

We then estimate the effect of variations in the value-based incentives on activity diversification of the NGOs using the Ugandan NGO data, conditional on a range

of organisational and managerial characters. To construct the variations in the incentives, we exploit between-NGO variations in the reliance on international aid and relief support after a historic, unexpected flood in the eastern and northern districts of Uganda from July to September 2007. We find an NGO that worked in the most affected districts during 2007 has a significantly higher proportion of contractual revenues (international donors, membership and user fees) than an NGO that worked in the less affected areas identified by the international aid community and the Ugandan Red Cross. Using a binary variable of whether the NGO worked in the most affected areas as the instrument for variations in incentives across NGOs, we show that NGOs receiving higher contractual incentives will diversify less than they would have done otherwise. To ensure the validity of our instrument (the location decision of the NGO before 2007), we rely on two assumptions: (i) having worked in the areas that were most affected by the flood affects the diversification decision only through the variations international aid relief and monetary incentives towards the most affected districts, (ii) the timing of the 2007 flood is exogenous. Using historical data on Ugandan rainfall, we show the flood was indeed an exogenous event (happened once in the last 100 years) and rainfalls in Ugandan districts follow a white noise pattern (Bjorkman-Nvqvist, 2013). We also demonstrate that the location decision (the instrument) is uncorrelated with any observable organisational and managerial factors that may affect diversification. To address the concern of potential unobservable pre-selection into locations vulnerable to extreme floods, we use the Dartmouth Flood Observatory global archive to construct a measure for flood vulnerability (that is, whether the NGOs worked in areas prone to at least one recorded extreme flood in the archive from 1988 – 2017). Our results remain when we restrict the analysis to NGOs who have operated in the most vulnerable areas, treating the 2007 flood as the source of exogeneity. The results rule out the concern that unobservable pre-selection of the location before 2007 drives our results.

Using an alternative strategy exploiting within-organisation variations between 2002 and 2007, time fixed effects and several time-varying characteristics, we obtain a similar result: NGOs respond to a higher reliance on incomes from contracts by focusing on fewer charitable activities. Using time fixed effects, we show during 2002 and 2007 these NGOs tend to diversify more. Combining with our model, the findings from both the within- and between-NGO variations suggest that the effects of the motivations to reduce funding-related uncertainty on diversification outweigh the effects of motivation to derive personal gains. Over time, NGOs diversify more once they receive financial stability from contractual sources. This highlights the importance of financial commitments and stability in ensuring NGOs to concentrate on overarching missions instead of actively expanding.

We contribute to the literature in three dimensions. First, our paper identifies the consequences of stakeholder funding on NGO behaviours. In combing two bodies of the

literature: the determinants of firm diversification (Campa and Kedia, 2002; Aggarwal and Samwick, 2003) and studies on expansion strategies of non-profits (Mendoza-Abarca and Gras, 2017), we contribute to the literature on behavioural motivators in organisational settings (Carpenter and Gong, 2016) and philanthropy (Tonin and Vlassopoulos, 2015). We also follow an agency approach to model the relationship between a stakeholder and an NGO. However, we differ in three overlapping dimensions. First, the previous literature focuses primarily on financial stability (Kistruck et al., 2013) as a reason for firm-level diversification. Instead, we consider additional potential motivations for diversification, namely, private gains of diversifying. There are also studies in the for-profit literature with analyses between product and stock portfolio diversification and the firm performance and efficiency, treating measures of diversification as explanatory terms (Arikan and Stulz, 2016; Carroll and Stater, 2009). We, instead, aim to explain the diversification decision. Second, the majority of the existing literature, including Aggarwal and Samwick (2003), rely on nonrandom datasets of corporations or international, large non-profits in developed markets, and use fixed effects or sample correction models to account for endogeneity (see Erel et al., 2015 and Aggarwal and Samwick, 2003). We differ with a unique, randomly sampled dataset of Ugandan NGOs, and use the exogeneity of the historic flood in Uganda in 2007 to identify the impact of contractual incomes on NGOs' diversification.

Second, we add to the existing development literature on motivations of prosocial behaviours by providing a new angle, diversification, to the agency approach in modelling an NGO's decision (see Steinberg, 2010 for applications in the non-profit literature). Our model relates to theoretical studies on the motives underlying decisions of non-profit organisations who often claim their acts as "voluntary" and driven by "intrinsic motivations" (Besley and Ghatak, 2005; Bénabou and Tirole, 2005). Fruttero and Gauri (2005) propose a model of location decisions of NGOs, where NGOs act strategically in response to donors by labelling projects as success or failure according to given measurable outcomes. The authors show that NGOs avoid even the neediest communities if the locations are excessively challenging. Barr and Fafchamps (2006) provide a similar result in the Ugandan NGO sector, suggesting that to avoid challenging environments, NGOs tend to cluster together, leaving the neediest communities without sufficient assistance. Instead of looking at the location decision of NGOs, we highlight the decision to diversify as a better way to understand what motivates NGOs.

Finally, our paper relates to experimental studies on designing incentives to motivate charitable efforts (see Gneezy et al., 2011 for early studies). Recently, Imas (2014) shows participants work harder for a charity than for themselves only when incentive stakes are low. DellaVigna and Pope (2017) experimentally show monetary incentives work far more effectively than psychological motivators. We emphasise the complementarity as our

paper also finds evidence for pragmatic incentives for reducing risks and enhancing survivability being dominant motivators rather than personal reasons such as altruism, prestige or signalling. It is a useful result for stakeholders when designing aid packages to incentivise their subcontractors, particularly in the development context where grassroots organisations often lack financial stability.

The paper proceeds as follows. Section 4.2 reviews the benefits and downsides of diversification. Section 4.3 outlines our theoretical predictions, which are tested by two empirical strategies in Section 4.4. Section 4.5 concludes with policy recommendations. Appendix for Chapter 4 provides a battery of robustness checks.

4.2. Why do development NGOs diversify?

4.2.1. Diversification to reduce risks related to funding

A popular result in the literature is that firms diversify both to mitigate against idiosyncratic uncertainty in their funding streams and to protect themselves from the ambiguity surrounding performance measures against which their effort is evaluated[58] (Montgomery, 1994). Since for-profit firms and charitable organisations have become increasingly similar regarding organisational struggles, managerial incentives, and resource scarcity (Boris and Steuerle, 2006), it becomes highly relevant for non-profit managers to engage in various income-generating activities to hedge themselves and to find the niche areas that they would fit. While NGOs in developed markets benefit from multiple streams of revenue (commercialised incomes, insurance, contracts, endowments) and donations that are readily available (Trussel and Greenlee, 2004); grassroots development NGOs have far fewer choices. Hodge and Piccolo (2005), Barr et al. (2004) document that grassroots NGOs mainly obtain funding from international donors, local development contributions or fees from the beneficiaries. As a result, any change in individual donor preferences or funding focus would lead to unexpected financial downturns for these organisations. NGOs with a narrow set of activities would struggle to adapt to new funding requirements if their single mission ceases to be targeted by funding bodies. To reduce dependence on a single source of income and the risk of interrupted funding, development NGOs could pursue a diversification strategy and engage in different activities or missions. Kistruck et al. (2013) and Keating et al. (2005) provide descriptive evidence of how diversification could provide a safety net for development NGOs and improve their financial stability. Carroll and Stater (2009) echo the result with a large sample of American non-profit organisations.

[58] An observable measure of the organisation's performance. In our context, it could be the number of people that an NGO was able to help during a humanitarian campaign.

In addition, since the performance of non-profits is usually difficult to measure, an appropriately diversified portfolio of activities could be considered as a "prudent" approach to harmonise the impact of not accomplishing one activity (Young, 2009, p.34). Carroll and Stater (2009) record cases in which NGOs providing higher quality services do not manage to attract increased funding. Scott (2003) shows that donors have begun to shift from measuring the success of development activities by anecdotal stories to measurements of accomplishments. He argues that having a broad portfolio could be a comparative advantage and an easily visible measurement of efficiency. Martens (2005) echoes the argument that due to the non-distribution constraint NGOs could signal their superior value to for-profit counterparts by maximising the number of activities until their entire given budget is spent. Empirical evidence has been consistent with the arguments of Scott and Martens (Kistruck et al., 2013; Carroll and Stater, 2009).

4.2.2. Diversification for private gains

The agency theory in the corporate literature proposes several managerial benefits and downsides of diversification (see Aggarwal and Samwick, 2003 for a related review). First, managers may wish to signal competence in order to improve their own future career prospects through their experience in managing complex NGOs. Evidence for such desire has been mainly drawn from the finance literature (Gibbons and Murphy, 1992). Indeed, having a well-diversified set of activities could increase the perception of the NGOs capability to the funding market since it signals experience and managerial ability. Perceived competence is particularly important for those who wish to tap into new funding. Through experiments, Aaker et al. (2010) replicate Moscarani (2007) to show that cues of credibility, if given, can serve as an effective tool to improve perceptions of competence, and thereby the likelihood of donations. Since performance remains difficult to verify and the number of activities offered is visible, an NGO's best strategy to enhance its credibility is thus to "vigorously" pursue a wide-ranging set of activities (Martens, 2005, p.650). Olennick (1988) reports that institutional donors tend to avoid risks by funding well-managed non-profits with rich and established portfolios, even when smaller firms exist that could yield a greater development value.

Second, diversification could also be associated with classic rent-seeking and entrenchment. Managers may wish to be involved in various activities in order to increase their potential earnings (Murphy, 1999). A related private gain is "managerial entrenchment" (Shleifer and Vishny, 1989, p.137). An agent may engage in activities that they are uniquely capable in the hope of increasing both the principal's demand for their skills and the cost to the donor of disposing of them. For NGOs, particularly small and grassroots organisations, the desire to maintain a relationship with donors is important

because of the fierce competition for funding (Aldeshev and Verdier, 2015). As a result, NGOs may diversify to entrench themselves when other organisations could become viable replacements, or when they are applying for new grants. In line with the risk reduction approach, diversification could mitigate the employment risk. Burger and Owens (2013) provide evidence that entrenchment concerns may well exist in the non-profit world. They find that grant-giving behaviour may be strongly habitual as the grant approval in the past raises the success rate in the subsequent round. The desire to prolong the relationship could thereby induce NGOs and managers to diversify.

Third, agents can derive egoistic motivations that include prestige, privilege and improved social status from running a more diversified organisation. Stulz (1990) analyses a model that incorporates financing policies to suggest that managers derive perquisites from expanding firms' portfolios. Egoistic gains, particularly prestige and recognition by others, also play an important role in philanthropic activities (Olson, 2009; Benabou and Tirole, 2006). An NGO that manages to provide a wide range of missions could be highly regarded by peers, particularly those belonging to international networks or local umbrella organisations. Roll (1986) even suggests the hubris hypothesis behind decisions of agents to diversify. He cites the psychology literature to argue that agents become overconfident in their capability and then more inclined, whether rationally or not, to improve social prestige by diversifying. Morfit (2011) finds evidence that Malawian NGOs with lower survival rates are those who become self-persuaded to take on AIDS-fighting campaigns regardless of their current skills. Pursuing social recognition in such a manner could induce non-profits to diversify.

The fourth source of private benefit possibly derived from performing more charitable acts is the "warm glow" feeling or impure altruism (Andreoni, 1989). Different from corporations whose objective is to maximise profits, NGOs could attempt to offer an extended set of activities to accommodate a wider range of beneficiaries. As there does not exist a generally accepted measure to study unobservable characteristics of altruism, the literature has not been able to reach a decisive conclusion for whether and how altruism affects the number of NGOs' activities. Edwards and Hulme (1992) find that despite inherent concerns of diversification, an expansionist strategy remains desirable to both multilateral assistance donors and NGOs. The main motivation cited by Edwards and Hulme is that operational expansion permits the organisations to benefit more people. Recent evidence from experimental economics, however, indicates that altruism may not translate into a desire to see more activities. Small et al. (2007) demonstrates that individuals donate the same amount for helping one person as for helping 10 people. Linardi and McConnell (2011) find that given excuses to stop volunteering, participants in a non-profit operation do not maximise the amount of altruistic effort that they are exerting.

Their study is consistent with Hsee and Rottenstreich's (2004) finding that as long as participants are working toward an altruistic activity, they may perceive the "warm glow" derived from the act independent of the number of donations and the number of activities. This new strand of literature implies that altruism does not always motivate non-profits and their managers to extend their portfolios.

Finally, diversification can incur costs and downsides. For one, the managers can experience management fatigue due to an increased workload of managing extra projects (see Gronbjerg, 1991 for four case studies). A greater variety of activities also leads to a corresponding increase in funding conditions and implementation criteria and satisfying one condition may preclude satisfying another. Besides adding to administrative overhead and goal conflicts, diversification may create confusion about the extent of mission accomplishment since different stakeholders may hold different views of effectiveness and goals (Tuckman, 1998). Ultimately, the public may perceive a non-profit chasing multiple roles and missions as having an unclear charitable purpose and "mission vagueness" (Weisbrod, 1998, p. 171). Such a public image may blur the distinction between the non-profit sector and their for-profit counterparts, fuelling the controversies surrounding the legitimacy of NGOs (Ferris & Graddy, 1989). Such costs may outweigh potential gains.

4.3. A conceptual model on incentives and diversification

We aim to model a contractual relationship between two players (an NGO and its stakeholder) that accommodates both the motivation to reduce risk and to benefit privately from diversification. We follow the spirit of Aggarwal and Samwick (2003) but differ in several aspects.[59] First, we assume that the contractible performance (development value) of the NGO is linear in the NGO's (unobservable) effort only. We do not assume diversification is contractible in order to relax a common assumption in the corporate literature that stakeholders discount diversification from the organisation's value (see Aggarwal and Samwick, 2003 for a review). In the corporate literature, when a firm's portfolio size increases their share price goes down. Whether stakeholders reward/punish NGOs who diversify remains a debate in the development context. By contracting on one performance measure (the development value), our model of a single effort level is equivalent to a dynamic model in a continuous time frame that allows the agent to dynamically adjust their effort over time (Holstrom and Milgrom, 1987). As such, we account for dynamic strategic decisions over time taken by the stakeholder and the NGO.

[59] We follow the modelling approach of Rose-Ackerman (1982) which assumes NGOs act as utility maximisers. We capture the non-distributional constraint by the condition that in equilibrium, the NGO would receive zero pay-offs for their choice.

Second, we explicitly model the personal benefits and costs of pursuing multiple missions in the NGO's utility function. In doing so, we relax another assumption in Aggarwal & Samwick (2003)'s framework that diversification incurs no costs. Although their assumption appears reasonable in the context of managing a diverse portfolio of many stocks and shares, the assumption does not apply to our context as managing different charitable activities even within one mission requires different levels of skills.

Third, we introduce a term representing the stakeholder's empathy toward the NGO to reflect the altruism of the stakeholder. Previous literature captures the shared intrinsic motivation between a non-profit and their funder (Prendergast, 2008; Besley and Ghatak, 2005). However, these papers explore incentives in mission-related production when the principal can use nonpecuniary interest in the organisation's success to incentivise the agent. In our model, we follow Stark and Falk (1988) to model the stakeholder's interest in the agent's welfare by including an additive term of empathy to the stakeholder's utility function. Finally, our focus is to derive comparative statics for the total effect of the incentives on the NGO's optimal diversification level, allowing exogenous parameters (defined below) to vary.

4.3.1. Building blocks

A risk-averse and effort-averse NGO (the agent) receives a mission-related payment from a risk-neutral and altruistic stakeholder (principal) in the form of grants, membership or user fees to deliver a charitable mission.[60] The agent chooses the level of diversification of their charitable activities $n \geq 1$, and unobservable effort $e \geq 0$, conditional on accepting the contract. We can think of the mission as assisting a depleted village, and the diversification as the number of charitable activities that the NGO proposes to undertake to assist the villagers. The activities can be wide-ranging. For example, the Ugandan Water and Sanitation Network's mission is to "alleviate poverty [...] through universal access to safe, sustainable water and improved sanitation".[61] Nevertheless, their activities in 2002

[60] Due to the nature of a contract, we exclude cases when the NGO receives a donation, which usually comes as a gift or contribution without any contractual requirements or strings attached. In this case, it is up to the NGO to use the donation or may disburse the donation as the NGO deems appropriate. For both theoretical and empirical modelling, we refer to grants, membership and user fees as mission-related payments as these forms of funding are contractible. Other revenue sources include donations, profits on special fundraising events, income from business (such as farming, canteens and retail outlets), incomes from services rendered to government, tax refunds and property/endowment incomes (Barr et al., 2005).

[61] http://www.uwasnet.org/Elgg/ [Accessed on 20 December 2017]

included advocacy and human rights, a library service, raising awareness, technical assistance, and community development. The NGO chooses (e, n) to maximise a payoff defined below.

The stakeholder's objective is to maximise the development value derived from the funded mission.[62]

$$v = e + \epsilon(n) \tag{1}$$

We can think of the observable development value as, for example, the number of villagers who were assisted following the contracted mission. This development value is made up of e which is the NGOs' observable charitable effort and is positively correlated with the outcome; and a normally distributed uncertainty $\varepsilon(n)$ which has a zero mean and a variance of σ^2/n. The variance captures two types of uncertainty. First, the ultimate value of the project could be affected by shocks specific to the organisation. Following the literature on diversification, an organisation with a more diverse portfolio of activities would be in a better position to reduce the impact of shocks that affected one particular activity. This assumption is reflected in the shock variance which decreases when $n > 1$. Second, since the performance of a mission remains difficult to measure (for example, due to being entwined with other ongoing projects), the variance captures the noisiness of the performance measure v. Following Martens (2005: p.650), we assume that having multiple activities could increase the precision of the performance measure. We refer to the variance σ^2/n as uncertainty.

Since monitoring the unobservable effort e is costly, the mission-related payment is based on the noisy signal of the development value v.

$$w = w_0 + \alpha v \tag{2}$$

where the NGO receives a lump sum of w_0 and a value-based component of αv. If the agent chooses (e, n), their overall wealth W increases by a monetary amount of w. In this setting, the fixed component w_0 can be interpreted as the NGO's endowment, such as equipment and vehicles, and is not taken into consideration in the ultimate development value. Through an NGOs grant application, the stakeholder is aware of the constraints faced by the NGO in the implementation of the project. Based on these constraints the stakeholder provides an incentive to the NGO to increase their effort α. Following Carroll (2015), we assume the incentive will be proportional to the realised development value.

[62] Although several rigorous conditions are required for a linear payment scheme to be truly optimal (Holmstrom and Milgrom, 1987), such a scheme is often used in practice for simplicity and relative robustness against manipulation (Dixit, 2002).

In addition to the payment w, the NGO derives net private benefit from choosing a diversification level n of $p(n) = \lambda(\ln n - \frac{n}{\tau})$ and incurs a disutility from exerting effort e of $K(e) = k\frac{e^2}{2}$. Parameters λ and τ respectively represent the marginal private benefit of diversification and the NGO's optimal diversification level in the absence of a contract. As in Section 4.2, private benefits may consist of prestige from managing a diversified organisation, rent-seeking motives, entrenchment, or the warm glow from helping others (Andreoni, 1989). Through this functional choice, we aim to reflect the reality that while diversifying could benefit the organisation when the number of activities remain manageable, over-diversifying ($n > \tau$) could be harmful to the NGO due to either inefficiency (Moore, 2000; and Martens, 2005) or management fatigue. Overall, the net private benefit associated with diversification increases when $n \leq \frac{\lambda}{\tau}$ and decreases when $n > \frac{\lambda}{\tau}$, where $\beta = \frac{\lambda}{\tau}$ is the marginal cost of diversification (effort fatigue and time due to managing multiple missions).

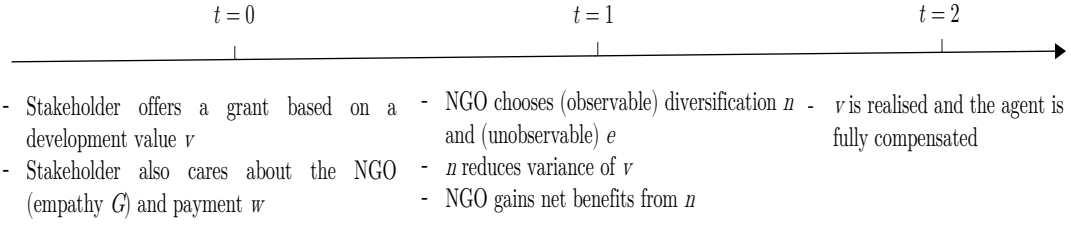
The disutility from the charitable mission (k) involves challenges and difficulties arising from current external environmental factors, such as poor infrastructure due to a natural disaster or the remoteness of an area. In our context, we can think of κ as a proxy for the conditions where the NGO is contracted to work – where the beneficiaries are located. As is standard in the literature, we assume $k > 0$ to reflect an increasing marginal disutility. We assume (k) is large enough for the NGO to apply for external funding from the stakeholder.

The induced wealth of the NGO for taking the contract is:

$$\varpi = w + p(n) - K(e) = w_0 + \alpha[e + \varepsilon(n)] + \lambda \ln n - \beta n - \frac{ke^2}{2} \tag{3}$$

To determine the certainty equivalent (CE) for taking the contract, we assume that the NGO has a constant absolute risk aversion utility function (CARA): $u(\varpi) = -\exp(-r\varpi)$ where r is a Pratt-Arrow absolute risk aversion coefficient. Like $\sigma^2, \lambda, \beta, k$, we assume r is observable before the interaction. Figure 4.1 summarises our three-stage model. The stakeholder maximises the development value net of the grant, given that the NGO chooses the effort and the diversification to maximise the NGO's utility.

Figure 4.1. Timeline for the model



4.3.2. Optimal actions and comparative statistics

We solve the game by induction. In the second period, the agent accepts the contract and chooses (e, n) to maximise their certainty equivalent, given the first-period incentives α and their own participation constraint.

$$\max_{e,n} CE = w_0 + \alpha + \lambda \ln n - \beta n - \frac{ke^2}{2} - \frac{r}{2}\alpha^2 \frac{\sigma^2}{n} \tag{4}$$

$$s.t \qquad CE = w_0 + \alpha e + \lambda \ln n - \beta n - \frac{ke^2}{2} - \frac{r}{2}\alpha^2 \frac{\sigma^2}{n} \geq u_0 = 0 \tag{5}$$

where u_0 is the reservation utility if the agent declines the contract in stage 1. To reflect the non-distributional constraint in the non-profit sector, we assume $u_0 = 0$. That is, the participation constraint holds when the NGO gains no payoffs (profit) at equilibrium.

Differentiating (4) with respect to e and n respectively gives the agent's optimal choice of charitable effort e^* and diversification amount n^*.

Corollary 1. The optimal amount of charitable effort and diversification are unique at:

$$e^* = \frac{\alpha}{k} \tag{6}$$

$$n^* = \frac{\lambda + \Delta}{2\beta} \tag{7}$$

where $\Delta = \sqrt{\lambda^2 + 2\beta r \sigma^2 \alpha^2}$.

The NGO's optimal effort increases in the incentive given by the stakeholders and decreases in the disutility (challenges) from exerting effort. This prediction is evident in Uganda and Bangladesh where NGOs are less likely to work in areas with most needs as they are often remote and challenging (Barr and Fafchamps, 2006; Fruttero and Gauri, 2005). Equation (7) provides an opposite prediction to evidence in the for-profit literature (Campa & Kedia, 2002): if the stakeholder provides more incentives, the agent chooses more diversification in response. Intuitively, it is consistent with the praise given to the nonprofit sector, namely that due to the non-distributional constraint these organisations tend to provide further services until their budgets are exhausted (Martens, 2005). However, we show below that this prediction does not hold if we allow for the stakeholder's incentive being chosen endogenously.

To calculate the optimal incentive package, the principal moves first and maximises their payoff, knowing the strategy of the NGO. We include two terms to capture the altruism of the stakeholder – the stakeholder's utility increases with the well-being of the beneficiaries and the contracted agents (Fehr & Schmidt, 2006): the expected development value $E(v)$ representing the value of the mission to the targeted beneficiaries, and the empathy $G(k, \alpha)$ towards the NGO. Following Stark and Falk (1998), the function of empathy towards the NGO $G(k, \alpha)$ captures the additional utility gained by the stakeholder from offering the contract to the NGO. Since the NGO gains zero payoff in

equilibrium $u_0 = 0$, we cannot model the empathy on utility gained by the NGO from accepting the contract. Instead, we model the stakeholder's empathy towards the NGO on two dimensions: the NGO's disutility of effort and the stakeholder's incentives given to the NGO. We assume the stakeholder derives a higher level of empathy toward the NGO with a higher disutility of effort, $\partial G/\partial k > 0$. The intuition is that a higher disutility of exerting effort is associated with a working environment where the beneficiaries require more assistance from the NGO. By definition, the stakeholder's empathy also increases with the offered incentive: $\partial G/\partial \alpha > 0$. The intuition is that for a higher incentive α, the NGO could derive a higher utility from exerting the same level of effort, thereby increasing the additional utility for the stakeholder's empathy. To allow for an explicit solution, we assume that $G(k, \alpha) = \alpha f(k)$ where $f(k)$ is a function such that $f'(k) > 0, f(0) = 0$. The first condition is consistent with the stakeholder gaining a higher utility from empathising with the NGO working in a more challenging environment. The second imposes the constraint that the stakeholder has zero empathy towards the NGO if the agent faces zero challenges in their working environment. We abstain from imposing any condition on $f''(k)$ for flexibility.

The final payoff of the stakeholder P_d is net of the contracted payment to the NGO. For simplicity, we follow Stark and Falk 's (1998) modelling approach and assume that the stakeholder's payoff is linear in their components. Using the fact that the participation constraint is now satisfied with equality or $w_0 + \alpha e - \beta n + \lambda \ln n - \frac{ke^2}{2} - \frac{r}{2}\alpha^2 \frac{\delta^2}{n} = 0$, the stakeholder's payoff can be written as:

$$P_d = E(v) + G(k, \alpha) - w = e^* + \alpha f(k) - \frac{k}{2}(e^*)^2 + \lambda \ln n^* - \beta n^* - \frac{r}{2}\alpha^2 \frac{\sigma^2}{n} \tag{8}$$

Knowing the response function (e^*, n^*) of the agent as in Corollary 1, the principal moves first and sets α to maximise their net expected payoff:

$$\max_{\alpha} P_d = \max_{\alpha} E(v) + G(k, \alpha) - w =$$
$$= \max_{\alpha} \frac{\alpha}{k} + \alpha f(k) - \frac{\alpha^2}{2k} + \lambda \ln\left(\frac{\lambda + \Delta}{2\beta}\right) - \beta \frac{\lambda + \Delta}{2\beta} - \frac{\beta r \alpha^2 \sigma^2}{\lambda + \Delta} \tag{9}$$

The first differential with respect to α following (10) is:

$$0 = \frac{1 - \alpha}{k} + f(k) - \frac{2\alpha\beta r\sigma^2}{\Delta + \lambda} \tag{10}$$

where $\Delta = \sqrt{\lambda^2 + 2\beta r \sigma^2 \alpha^2}$.

Lemma 1: There exists a unique value-based incentive $\alpha^* > 0$

Proof. To see this, let $\alpha \to 0$, the right-hand side of (10) becomes positive $(\frac{1}{k})$; while when $\alpha \to 1 + kf(k)$, the right-hand side of (10) is negative $(-\frac{2\beta r\sigma^2}{\Delta + \lambda})$. Therefore, there exists at least one interior solution on $(0, kf(k))$. It is trivial to see that the right-hand side is a

strictly decreasing function of α so the interior solution is also the unique solution for the first order condition. QED.

As expected, the agent is rewarded for delivering a higher development value $\alpha^* > 0$. In addition, solving α from Equation (10) implies that any exogenous determinants that affect the decision to diversify also affects the stakeholder's optimal value-based incentive α^*. Intuitively, grants given to NGOs are not determined exogenously, but jointly determined with factors associated with the NGO, namely, risk parameters and the working environment of the NGO. Substituting α into (7), the theory predicts the following structural form of the optimal level of diversification:

$$n^* = \frac{\lambda}{2\beta} + \frac{1}{2}\sqrt{\left(\frac{\lambda}{\beta}\right)^2 + 2\frac{r}{\beta}\sigma^2[\alpha^*(r, \sigma^2, \lambda, k, \beta)^2]} \tag{11}$$

Equation (11) suggests that any change in the diversification n^* due to a variation in the disutility k comes only through its effect on incentive α^*. In contrast, any variation in private benefits λ, personal cost of diversification β, risk aversion r, or the variance of shocks s would simultaneously affect both the diversification n^* and incentives α^*. Intuitively, since all parameters, except for the disutility from working k, are directly related to diversification, the stakeholder will exploit these relationships in setting the grant conditions. Corollary 2 provide a decomposition the correlations, highlighting how each individual factor affects the decision of the NGO.

Corollary 2: Given that working on a development project imposes a sufficiently high disutility of effort (k is high), the partial impact of the exogenous factors on the optimal level of diversification n^* and the optimal value-based incentive α^* are as follows:

$$\frac{\partial \alpha^*}{\partial j} < 0, j \in \{r, \sigma^2, \beta\} \qquad \text{and} \qquad \frac{\partial \alpha^*}{\partial k}, \frac{\partial \alpha^*}{\partial \lambda} > 0 \tag{12}$$

$$\frac{\partial n^*}{\partial j} > 0, j \in \{r, \sigma^2, \lambda, k\} \qquad \text{and} \qquad \frac{\partial n^*}{\partial \beta} < 0 \tag{13}$$

Proof: Differentiating (10) and (11) w.r.t $\{r, \sigma^2, \lambda, k, \beta\}$. See Appendix 4.6 for details.

The NGO's disutility of effort being high is a sufficient condition for Corollary 2. We argue that this assumption reflects the development context: working on development projects (missions), particularly in sub-Saharan Africa, is challenging and requires great dedication from aid workers. With that in mind, Array (12) illustrates how the optimal diversification level varies when each of the exogenous parameters varies, holding the others constant. When the parameters related to the uncertainty of performance measure (σ^2, r) decrease, the stakeholder increases the incentive because the performance can now be matched more precisely with the actual effort. This also results from the assumption that the NGO is risk averse. When the cost of diversifying decreases or the personal benefit of diversification increases, the stakeholder raises the incentive α^* to incentivise the NGO to

exert effort instead of diversifying. When the disutility of exerting effort k is higher, the stakeholder intuitively provides a higher incentive for the NGO to work on the mission. Array (13) shows the correlations between variations in the factors and the optimal diversification level. When the personal cost of diversification decreases or the personal benefit increases, the NGO diversifies more. Likewise, when the uncertainty of performance measure or the risk aversion parameter increase, the NGO switches to diversification to reduce the uncertainty. When the disutility of exerting effort is higher, the NGO diversifies more intuitively because it is costlier now to exert more effort. Mathematically, the increased disutility causes an increase in the optimal incentive (Equation 11), through which the NGO increases their diversification.

Proposition 1 links the partial differentials in Array (12) and (13) to show the impact of a change in the incentive on the optimal diversification level when each of the five factors $\{r, \sigma^2, \lambda, k, \beta\}$ varies, holding the other four fixed.

Proposition 1: The incentive effect on diversification depends on the five factors:

$$\left.\frac{\partial n}{\partial \alpha^*}\right|_{\partial r} = \frac{\partial n}{\partial r} \Big/ \frac{\partial \alpha^*}{\partial r} < 0 \qquad \left.\frac{\partial n}{\partial \alpha^*}\right|_{\partial \sigma^2} = \frac{\partial n}{\partial \sigma^2} \Big/ \frac{\partial \alpha^*}{\partial \sigma^2} < 0 \tag{14}$$

$$\left.\frac{\partial n}{\partial \alpha^*}\right|_{\partial k} = \frac{\partial n}{\partial k} \Big/ \frac{\partial \alpha^*}{\partial k} > 0 \qquad \left.\frac{\partial n}{\partial \alpha^*}\right|_{\partial \lambda} = \frac{\partial n}{\partial \lambda} \Big/ \frac{\partial \alpha^*}{\partial \lambda} > 0 \qquad \left.\frac{\partial n}{\partial \alpha^*}\right|_{\partial \beta} = \frac{\partial n}{\partial \beta} \Big/ \frac{\partial \alpha^*}{\partial \beta} > 0$$

Proposition 1 suggests that the magnitudes of the five exogenous parameters determine the sign of the overall effect of the stakeholder's incentive on the optimal level of diversification. Taking the total derivative of $\partial n / \partial \alpha^*$ from Equation (11), Proposition 2 shows how the sign of this derivative depends on the two components corresponding to the two explanations discussed in Section 4.2: to reduce risk or to derive personal gains.

Proposition 2. If the dominating motivation for a diversified portfolio is risk reduction, the total effect of a change in the incentive on diversification is negative. If the disutility and personal gain motivations are sufficiently high to offset the effect of the risk reduction motivation, and thereby determine the NGO's choice, the effect of a change in the incentive on diversification is positive.

$$\underbrace{\frac{\partial n}{\partial \alpha^*}}_{\substack{\text{Incentive effect of } \alpha \\ \text{on } n}} = \underbrace{\underbrace{\frac{\partial n}{\partial r} \Big/ \frac{\partial \alpha^*}{\partial r}}_{<0} + \underbrace{\frac{\partial n}{\partial \sigma^2} \Big/ \frac{\partial \alpha^*}{\partial \sigma^2}}_{<0}}_{\text{diversify for uncertainty reduction}} + \underbrace{\underbrace{\frac{\partial n}{\partial \lambda} \Big/ \frac{\partial \alpha^*}{\partial \lambda}}_{>0} + \underbrace{\frac{\partial n}{\partial k} \Big/ \frac{\partial \alpha^*}{\partial \kappa}}_{>0} + \underbrace{\frac{\partial n}{\partial \beta} \Big/ \frac{\partial \alpha^*}{\partial \beta}}_{>0}}_{\substack{\text{diversify for personal gains} \\ \text{(private benefits, costs or disutility)}}} \tag{15}$$

The implication of Proposition 2 is that if the NGO diversifies mainly to reduce the idiosyncratic uncertainty in their funding streams and to protect themselves from the ambiguity surrounding performance measures, they would respond to an increased incentive

from the stakeholder by diversifying less.[63] Whereas, if the NGO diversifies mainly to gain personally, the response to increased incentives will be to diversify more. The intuition is that if the NGO cares more about the uncertainty of the performance measure or the uncertainty related to funding in general, following a higher incentive they would exert more effort on core activities of the mission that the stakeholder aims to incentivise, instead of working on other activities. In contrast, if the NGO cares more about personal perks such as the benefits of diversification or the disutility of exerting effort, receiving higher incentives to undertake more charitable effort would instead allow the NGO to diversify more. The increased incentives now act as insurance for the NGO to offset risks related to the performance measure and future funding.

In summary, we extend an agency model to study the relationship between a stakeholder and a risk-averse NGO. Under the assumption that the NGO might diversify to reduce risk and to derive personal gain, we show that the effect of an increase in the incentives on the level of diversification is decomposed as below.

i. The incentive effect on diversification is negative if the underlying motivation for diversification is to reduce uncertainty related to performance measures or future funding. The magnitudes of either or both of r, σ^2 dominate the other parameters in determining the effect of incentives on the NGO's diversification level.

ii. The incentive effect on diversification is positive if the underlying motivation for diversification is for personal gain: to derive private benefit (λ), to avoid discomfort from working on projects (κ), or because the cost of diversification (β) is low. The magnitudes of either or both of λ, k, β dominate the other parameters in determining the effect of the incentives on the diversification level.

4.4. Empirical analysis

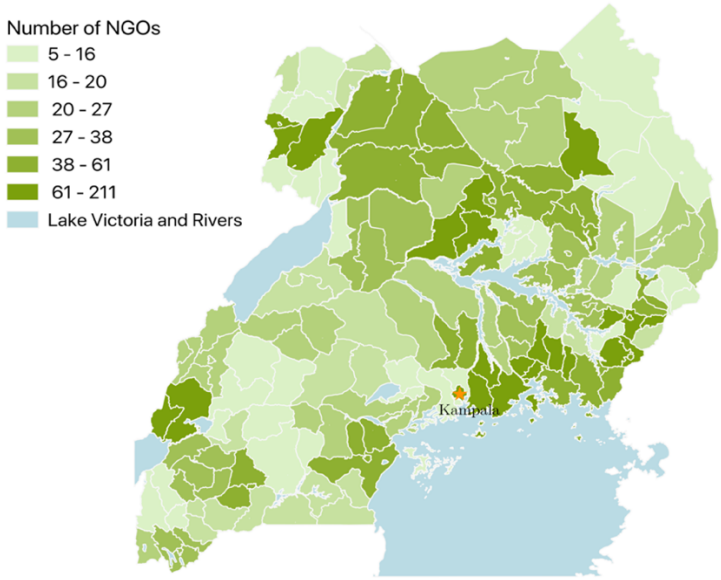
4.4.1. Data: the 2008 Ugandan NGO survey

We test the above predictions using a unique survey of the Ugandan NGO sector. Uganda is an interesting case to investigate the NGO behaviours due to its expanding NGO sector following initiatives by the international aid community and the Gates Foundation since 2000. increasing importance of the sector in distributing aid and delivering projects (Morfit, 2011). Furthermore, we are not aware of any other nationally representative dataset of the NGO sector available. We use the second wave of a nationally representative survey of the Ugandan NGO sector conducted in 2008, following the random sample collection protocol in Barr et al. (2005). Figure 4.2 shows the number of NGOs that either worked or had an

[63] Such as a sudden change in the government funding objectives, or a sudden shortage of labour to carry out projects.

office in each district of Uganda in 2007. By sampling design, the number of NGOs in each district in the survey reflects the pattern of NGOs registered in each district according the NGO Registration Board and verified by Barr et al (2005). The most popular locations were Kampala (211 NGOs), Abim (182 NGOs), Bududa (94) and Arua (86). The map also demonstrates that our sampled NGOs did indeed have a presence in all Ugandan districts, further supporting the representative nature of our survey.

Figure 4.2. The geographical coverage of NGOs in 2007

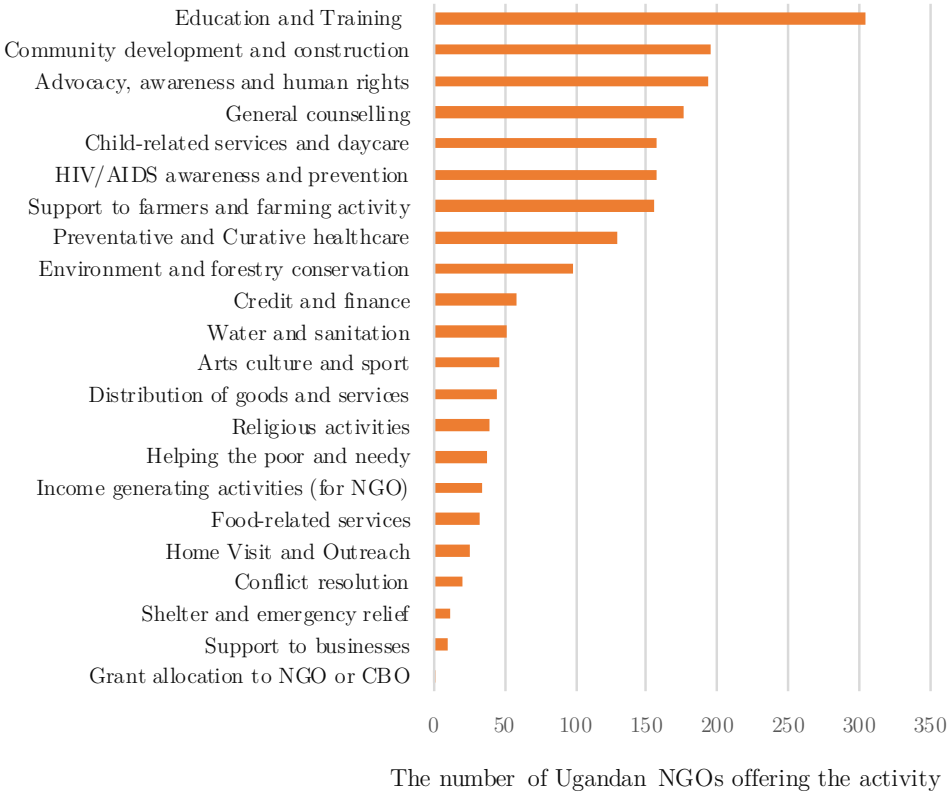


Notes. The map shows the number of NGOs working in each Ugandan district at the end of 2007. A darker colour represents a higher number. *Sources:* Authors' calculation using the 2008 Ugandan NGO survey by Burger et al. (2008). Table 4.A4 in Appendix 4.6 presents the detailed numbers for each district. The official spatial location is as of 2006.

The survey includes information about all of the activities undertaken and the proportions spent on each at the end of 2007. We also collect a wealth of information on other characteristics of the NGOs, such as their revenues from different funding sources, employment size, geographical coverage, governance structure and managers' background. From a total of 478 interviewed NGOs, due to missing information on one or several variables of interest, 391 NGOs are included in our analysis. Table 4.A5 in Appendix 4.6 demonstrates that there is no significant difference in observable characteristics between the 391 NGOs and the 87 NGOs excluded from our study.

Figure 4.3 presents the categorised activities and their prevalence in the Ugandan NGO sector at the end of 2007. The most popular activities in 2007 include Education and Training, Community Development, Advocacy and Human Rights, HIV/AIDS awareness and prevention, Support to farmers and farming activities, Child-related Services, Counselling, Water and Sanitation, Credit and Finance. Burger and Owens (2013) provide further details of the data and the sampling method.

Figure 4.3. The prevalence of categorised activities in 2007



Source: Authors' calculation using the 2008 Ugandan NGO survey by Burger et al. (2008).

4.4.2. Empirical strategy

Our objective is to identify the dominant motivation underlying an NGO's decision-making, in particular their decision to diversify. Following our theoretical framework, we do so by observing the responses of the NGO's diversification to a change in incentives set by their stakeholders. Section 4.3 presents two possible scenarios where the shifts in either risk reduction or deriving personal gains dominates an NGO's diversification decision. The model suggests that observing the sign of the overall effect of incentives on the observed level of diversification could indicate the underlying mechanism. Ideally, we would like to estimate the structural form specified in Equation (11), however, as the parameters $s^2, \lambda, k, r, \beta$ are not observed in our data, we instead identify the effect of interest in the following reduced form:

$$n_i^* = \gamma \text{INCENTIVES}_\text{i} + \text{X}_\text{i}'\gamma_\text{X} + \varepsilon_\text{i} \tag{16}$$

where γ is the parameter of interest: the average effect of an additional unit of incentive from the stakeholders on the level of diversification, X_i' is a set of control variables adopted from previous studies to mitigate the concern of omitted variables. Following Proposition 2, a positive estimate, $\hat{\gamma} > 0$, indicates that the magnitudes of the incentive effect caused by private-gain parameters λ (private benefits), k (disutility intolerance), and

β (private cost of diversification) dominates the partial incentive effect caused by risk-related parameters δ^2 (idiosyncratic uncertainty) and r (risk aversion). In contrast, a negative estimate, $\hat{\gamma} < 0$, indicates the magnitude of the effect caused by the risk-related parameters dominates the effect caused by the private gain parameters.

There are two main challenges for estimating the total effect of value-based incentives on the diversification level. First, there may be measurement error in NGOs' self-reported spending on activities. This measurement error could potentially cause downward attenuation bias of our estimates. However, as the error is unlikely to be correlated the pre-2007 location decision – our instrument (discussed below), our IV strategy would be able to alleviate this concern. Second, there is the issue of omitted variable bias. Equation (11) suggests that both the optimal diversification level and the incentives given by stakeholders are jointly determined by the five parameters $s^2, \lambda, k, r, \beta$. That is, NGOs who offer more activities could also attract more revenue from stakeholders due to unobserved characteristics which we identify through the model such as the level of risk-taking (r), the uncertainty surrounding performance measures and future funding (s^2), the disutility of the working conditions (k), and the unobserved cost and private benefit of diversification (β, λ). Failure to capture these characteristics will bias $\hat{\gamma}$, and even reverse the sign of the true estimate. In the next section, we describe our variables of interest and strategies to address this concern.

We use two measures of activity diversification n^*. First, we use the categorised activities specified in Figure 4.2 to construct a count index of the number of activities that each NGO reported to participate in at the end of 2007.[64] Second, we use a Herfindahl-Hirschman index (HHI) to measure the diversity of the organisation's expenditure over different categories of activities to address the concern that NGOs may be able to increase their activity diversification by simply listing activities whose allocated budgets were negligible. The construction of the HHI index is:

$$\text{HHI} \quad = 1 - \sum_{j=1}^{n} \eta_j^2 \tag{17}$$

where η_j is the reported budget share of activity j. If an NGO spends most of their budget on one activity, the index reaches its lower bound at zero: the NGO does not diversify. For the HHI index being one, the NGO either delivers infinite activities or allocates insignificant amounts of money to all activities. Like the count index, a higher HHI indicates a higher level of diversification.

[64] The count measure of activities captures the extent (range) of each organisation's portfolio. Our data do not permit capturing the intensity of the categories – an NGO offers an activity closely related to its current activities instead of a completely new service.

We prefer the count index for two reasons. First, it provides a direct test of our theory as it is potentially observable to the stakeholder and the condition that $n \geq 1$ in the theory. The interpretation of our model also revolves around the count of activities provided within a mission of NGOs. Second, we were only able to construct the HHI measure for 330 out of 402 organisations because 72 NGOs did not provide any financial information on how they allocated their spending across activities. As such, an analysis using the count index could maximise the available sample caused by the missing values.

To measure the value-based incentive INCENTIVES_i, which captures the contractual relationship of the NGO with its stakeholder(s), we use the income from donor grants (local and international), membership and user fees as a percentage of total income.[65] By construction, the measure captures the sensitivity of an organisation's income to the contractual relationship with their stakeholders.

To mitigate the omitted variable bias, we propose a set of control variables to proxy for the exogenous parameters. To capture heterogeneity at an organisational level, we include DISTRICTS that indicates the number of districts in which the NGO has staff working in 2007. A binary KAMPALA variable takes the value 1 if the NGO has its headquarters in the capital and 0 otherwise. Two binary variables GEOGRAPHICAL EXPANSION and CHANGE FOCUS take the value 1 if the NGO has expanded geographically or changed its focus in the last five years, respectively; 0 otherwise. VOTE_ACTIVITY takes the value 1 if the NGO requires a vote from either its oversight committee or its members or some external body before introducing a new activity; 0 otherwise. Finally, we control for the organisational size with LOGSTAFF - the logarithm of the number of staffs working for the NGO. To capture heterogeneity at managerial level, we include TENURE and TENURE2 indicating how long the current manager has been with the NGO and its square to capture the standard career concerns found in the literature and a possible U-shape relationship. To proxy for altruism, we use RELIGIOUS equal to 1 if the manager holds a religious title, 0 otherwise.[66] To proxy for time resources available for managing a diversified NGO, we include OTHER_NGOS indicating the number of other NGOs that the manager is currently involved.

---

[65] There are 85 organisations listing user fees as one of their sources of income. These organisations on average offer similar services to those without user fees. Main activities include Counselling, Education and Training and HIV/AIDS prevention. Four organisations report incomes solely from user fees.

[66] Adding a variable for female manager would reduce the F-stat for the excluded instruments test, but the estimates are unchanged.

Table 4.1. Descriptive statistics for Ugandan NGOs in 2008

VARIABLES	(1) Mean	(2) SD	(3) Min	(4) max
Number of activities	4.288	1.861	1	10
Herfindahl-Hirschman index (HHI)	0.538	0.224	0	0.993
INCENTIVES	61.59	40.27	0	100
KAMPALA	0.393	0.489	0	1
CHANGE FOCUS	0.242	0.429	0	1
DISTRICTS	3.750	5.986	1	57
GEOGRAPHICAL EXPANSION	0.467	0.500	0	1
NUMBER OF STAFF	34.87	94.01	1	1,284
VOTE_ACTIVITY	0.434	0.496	0	1
FEMALE	0.260	0.439	0	1
RELIGIOUS TITLE	0.227	0.419	0	1
TENURE	6.670	5.392	0.250	45
OTHER_NGOS	1.444	6.342	0	108
FLOOD_AFFECTED	0.196	0.398	0	1

Notes: INCENTIVES is measured as revenue from local grants, international grants, and membership and user fees as a percentage of total revenue. KAMPALA takes the value 1 if the NGO has the headquarter in the capital and 0 otherwise. GEOGRAPHICAL EXPANSION and CHANGED FOCUS take the value 1 if the NGO expands geographically or changes its focus in the last five years, respectively; 0 otherwise. VOTE_ACTIVITY takes the value 1 if the NGO requires a vote from either its oversight committee or its members or some external body before introducing a new activity. NUMBER OF STAFF is the number of staff working for the NGO in 2007. RELIGIOUS TITLE takes the value 1 if the manager has a religious title, 0 otherwise. TENURE indicates how long the current manager has been with the NGO. OTHER_NGOS indicates the number of other NGOs that the manager is currently involved. There was one NGO whose manager reports working for more than 100 NGOs. If we exclude this NGO, the results remain unchanged. See Table 4.A7 in Appendix 4.6 for a sensitivity analysis when we exclude NGOs whose managers working for more than 10, 15, and 100 NGOs. FLOOD_AFFECTED takes the value 1 if the NGO reported to have worked in a district affected by 2007 flood. There are 391 NGOs in the sample.

Table 4.1 presents the summary statistics of the key variables. On average, Ugandan NGOs in 2008 operate approximately four activities; five report to be involved in 10 different activities and 27 focus on one activity. Figure 4.A1 illustrates the frequencies of the activities, showing that the distribution largely resembles a normal distribution. Regarding incentives, the surveyed NGOs rely on contractual funding from stakeholders for on average 62% of their revenue – 117 NGOs depend exclusively on this funding source

while 52 receive their revenue only from donations or business income. 40% of the surveyed NGOs have their headquarters in Kampala and on average work in 4 districts. About 40% of the NGOs require a vote from their members to introduce a new activity. The Ugandan NGOs on average have 35 members of staff in 2007 (the median is 14). Regarding the management, 25% of the managers are female; 23% hold a religious title. On average, managers have worked for more than one organisation before their current tenure (averaged at 7 years). Over the 2002 – 2007 period, 49% have expanded their geographical coverage; and 23% have changed their focus.

Regardless of the richness of control variables, the concern of omitted variable bias remains. We address the endogeneity through an instrumental variable strategy. The valid instruments must strongly correlate with INCENTIVES, namely, the proportion of income from stakeholders, but be uncorrelated with the error terms ε_i. To satisfy the exclusion restriction, the source of identification should exogenously cause a change in the incentives, and only through the incentives it affects the diversification decision of the NGOs. One mechanism is that exogenous variations in the working conditions of charitable activities in a location would result in the stakeholder changing the incentives (as a compensation) in order to induce the NGO to undertake extra charitable effort and then change the diversification behaviour. If the source of the variations in the incentive remains orthogonal to other pre-determined characteristics of the working location and the NGOs, the excludability condition of our instrumentation would be valid.

Guided by the theoretical mechanism, an ideal source of exogenous variations in the incentives would be an unexpected natural event that temporally causes unforeseen hardship in working conditions, exogenously requiring a shift in the composition of revenues of the NGOs working in the location. Our conjecture is that the donor community would then respond to these unexpected hardships by raising their incentives (the weight of grants or aid reliefs to the NGOs' incomes) specifically for NGOs working in the affected areas. The NGOs would then respond to the change in the incentives by varying their activities portfolio. Proposition 2 provides interpretations for the sign of the effect, implying which of the theoretical parameters dominantly affect the behaviour.

With this mechanism in mind, we exploit the surge in international relief aid specific to several areas in Uganda following a historic flood in Africa during July and September 2007. Both the flood and the surge in international grants to the affected areas are plausibly exogenous to the diversification behaviours of the NGOs. According to the UN Office for the Coordination of Humanitarian Affairs (UNOCHA) and the Ugandan Red Cross, the unexpected heavy rainfall led to flooding and damages across several districts in eastern and northern Uganda. The United Nation reported the 2007 flood as one of the

worst floods in recorded history.[67] Figure 4.3 demonstrates the unexpectedness and the severeness of the event. Using monthly rainfall data from The Climate Change Knowledge Portal (Word Bank), Figure 4.3 shows that the rainfall record for July 2007 was above 250 mm/month, nearly two times higher than the second rainiest figure during the period 1901 to 2015, a once in a hundred-year event.[68] According to the Dartmouth Flood Observatory archive, the event caused nearly 520,000 people to be displaced and at least 52 casualties.[69]

Following the event, there was also a sudden rise in the humanitarian response from international donors and aid institutions, particularly from the UNOCHA and the Ugandan Red Cross.[70] Exploiting this surge of international aid towards districts affected by the event, we conjecture that NGOs who had a presence in the affected districts prior to the flood would see a dramatic change in their income composition in 2007: contractual revenues would become a larger part of these NGOs' total income. There are two reasons. First, the surge in revenue from international aid during 2007 would allow organisations operating in the affected areas to shift focus from generating their own income towards receiving the (unexpected) funding from the new sources. Second, income from non-contractual sources such as local donations or business income is likely to decrease following the flood as local inhabitants have fewer resources available for donating or purchasing services. To the extent that the flood timing is as good as random, we instrument for INCENTIVES by a variable FLOOD_AFFECTED, which takes the value 1 if an NGO had worked (having staff or an office) in the most affected areas identified in the reports of the UNOCHA and the Ugandan Red Cross before 2007. The official list includes Soroti, Amuria, Katakwi, Bukedea, Kumi, Lira, Sironko, Bududa and Nebbi (see Figure 4.A2 in

---

[67] http://news.bbc.co.uk/1/hi/world/africa/6994995.stm [Accessed on 20 December 2017]

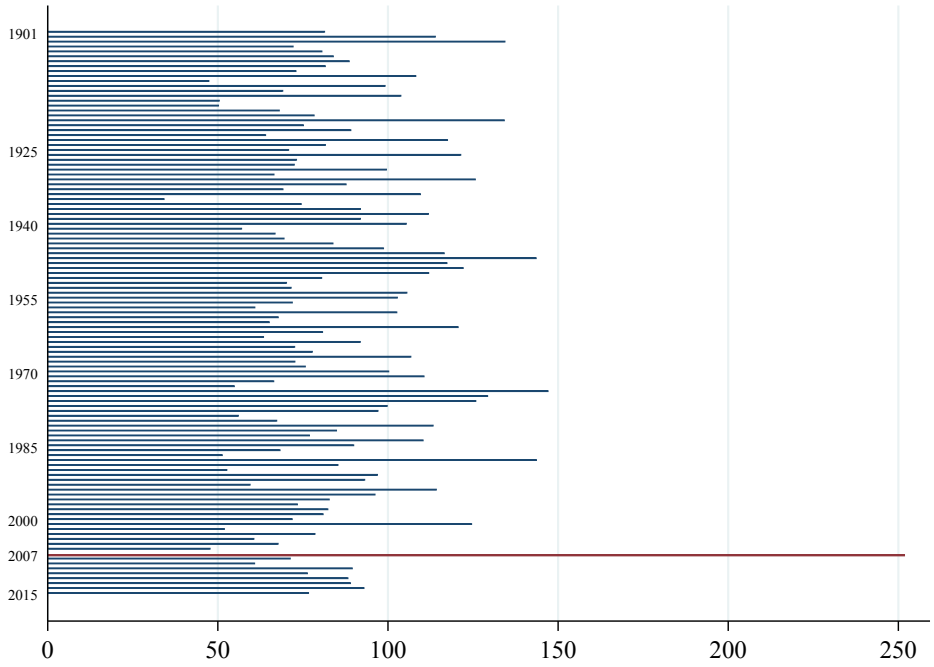
[68] http://sdwebx.worldbank.org/climateportal/ [Accessed on 20 December 2017]

[69] The 2007 flood was assessed at 1.5 on a 1-2 severity scale. There are three classes. Class 1 includes large flood events with significant damage to structures or agriculture. Class 1.5 includes very large events: with an estimated recurrence interval greater than 2 decades but less than 100-years, and/or a local recurrence interval of 1-2 decades and affecting a large geographic region (> 5000 sq. km). Class 2 includes extreme events with an estimated recurrence interval greater than 100 years. To highlight the uniqueness of the flood, the BBC's profile of Uganda documents the 2007 flood as the only weather event mentioned in the country chronology (mostly conflicts and political incidents).
Sources: https://www.dartmouth.edu/~floods/Archives/ArchiveNotes.html]
http://www.bbc.co.uk/news/world-africa-14112446

[70] [Accessed] March 2018] http://reliefweb.int/disaster/fl-2007-000138-uga (UNOCHA) and http://www.ifrc.org/docs/appeals/07/MDRUG006.pdf (Ugandan Red Cross)]

Appendix 4.6 for a graphical demonstration). Table 4.1 reports that 20% of the NGOs fall into this category. We expect that having a presence in the affected areas prior to 2007 would be strongly positively correlated with the endogenous variable INCENTIVES: NGOs in these areas would have a significantly higher proportion of their revenue coming from grants, membership and user fees.

Figure 4.4. The average rainfall (mm) in July from 1901 to 2015



Notes: The rainfall record for July 2007 was above 250 mm/month, nearly two times higher than the second rainiest figure during the period from 1901 to 2015. Sources: Authors' calculation using rainfall data from The Climate Change Knowledge Portal (Word Bank).

Table 4.2 strongly supports this expectation. Having a presence in an affected area significantly increases the proportion of contractual revenue from stakeholders (international, local grants and from user fees during the aftermath).[71] The Sanderson-Windmeijer F-statistic is large at 31.70, supporting for our instrument being strongly correlated with INCENTIVES. The Kleibergen-Paap rk LM statistic (22.91) rejects the null that the specification is underidentified. In case our instrument is only weakly

[71] We further support the first-stage mechanism in Table 4.A7 in Appendice. Using data on the net official development assistance (ODA) and aid received (constant 2013 US dollar) in 1960-2015 from the World Bank, we show that aid flows to Uganda are highly correlated with monthly average rainfall. Despite being descriptive, the result addresses the concern that the aid surge in 2007 might not be due to the flood but some political cycle that coincided with the flood timing. If this were the case, our exclusion restriction would have been violated.

correlated with INCENTIVES despite these test statistics, we additionally report Anderson-Rubin confidence intervals in Table 4.3 (see Cameron and Miller, 2015; Asatryan et al., 2017 for examples). We calculate bootstrapped confidence intervals with 200 replications to address potential failures of the underlying distributions. If these confidence intervals do not include zero points, it strongly suggests that our instruments have captured a significant and robust relationship even under the presence of weak instrumentation. To preview, the result from the Anderson-Rubin and bootstrapped confidence intervals at 200 replications supports the result from our IV estimations.

Our strategy of using the flood instrumentation relies on the assumption that, conditional on the controls, being affected by the unexpected flood in 2007 only affects the level of diversification in 2007, by causing an increase in the weight of incomes from international grants in the total revenues of NGOs working in the affected districts. Also, the cause for the surge in relief aid towards these areas is that the donors wish to compensate and motivate the NGOs for working in these tougher environments. The main concern with this assumption is that the allocation of NGOs to a different location is *ex-ante* not random: some NGOs could self-select into areas that are more prone to future floods in the hope of receiving consequential aid relief and the self-selection is correlated with the behaviours of the NGOs. To violate this exclusion restriction, NGOs operating in affected and unaffected areas would need to have systematically different characteristics. We discuss the excludability of our instrument in Section 4 and provide in Section 5 an alternative strategy using within-NGO variations.

Table 4.2 First stage estimation for the instrument $FLOOD_AFFECTED_i$

INSTRUMENTAL VARIABLES	*Dependent variable:* INCENTIVES
FLOOD_AFFECTED	25.39 ***
	(4.65)
Sanderson-Windmeijer multivariate F test of excluded instruments: (Prob > F)	31.70 ***
	(0.00)
Kleibergen-Paap rk LM statistic (underidentification) (p-val)	22.91***
	(0.00)

Note: *** p < 0.01. Robust standard error in parentheses. The first-stage estimation is:

$$INCENTIVES_i = \tau_1 FLOOD_AFFECTED_i + X_i'\tau + u_i$$

where X_i' is the list of control variables; u_i is an error term. The Sanderson-Windmeijer F-test statistics is equivalent to usual F-statistics used to compare against Stock-Yogo weak ID test critical value. The null of Kleibergen-Paap test is the model is under-identified.

As with all instrumental variable estimates, our 2SLS estimates reflect the average effect for observations that comply with the instrument, i.e., a local average treatment

effect (Imbens and Angrist, 1994). In our context, compliers are organisations that receive relatively more international relief aid following the 2007 flood. Our 2SLS estimates are not driven by the effect of incentives from stakeholders for organisations whose composition of revenues is unaffected following the flood.

4.4.3. Empirical results

Table 4.3 presents our main findings using our preferred measure of activity diversification, the count index. Ideally, to account for potential correlations between NGOs within districts, we would report estimates with standard errors clustered at the district level where the NGO's head office is. However, as the number of clusters is small (22), we report robust standard errors in the main analysis. Appendix 4.6 reports confidence intervals with pairwise bootstrap-based clustered errors as suggested by Ibragimov and Muller (2010) and Cameron et al. (2008). Overall, our results are not sensitive to clustering standard errors.

Column 1 reports estimates of the correlation between the reliance on income from stakeholders (INCENTIVES) and the level of diversification measured by the naïve count of activities. In a univariate estimation, there is a positive but statistically insignificant correlation. The estimates in Column 2 show that the univariate OLS relationship between diversification and incentives is unaffected by the inclusion of control variables. We find similarly small and statistically insignificant estimates. If we take these estimates as they are, the implication from our theory would be that the effect of shifts in risk-related parameters on the diversification behaviour is outweighed by the effect of shifts in parameters related to personal gains of diversification. That is, Ugandan NGOs, on average, diversify more in order to mainly derive personal gains.

As explained in sub-Section 2, the above estimates could be biased and inconsistent due to omitted variable bias. Indeed, once we control for potential endogeneity using the instrumental variable strategy, the overall incentive effect becomes negative and statistically significant.[72] Column 3 reports 2SLS estimates using the flood instrumentation. Like the OLS estimation, the 2SLS estimates remain similar as we introduce the controls (without control estimates are unreported to save space). Column 4 reports consistent results for the IV-Poisson estimates which deal with the count nature of the dependent variable. Larger reliance on contractual incomes from stakeholders leads to a more focused

[72] Although the magnitudes between the base and the IV estimates are different, it does not point out the invalidity of our instrumentation. One explanation is that the base estimates are over the whole population of the Ugandan NGO sector, whereas our IV estimates are for the subgroups whose behaviours are influenced by the instrument (the variations in incentives due to the 2007 flood).

portfolio of activities, suggesting that the motivation to reduce idiosyncratic risk (relating to the risk aversion of the organisation leadership and the external volatility faced by the NGO) dominates the motivation of deriving personal gain (including private benefits and disutility from undertaking activities). According to the estimates using the full set of controls, a 40% increase in the dependence of the organisation's revenue on stakeholder's grants, membership, and user fees decreases the number of activities offered by one (0.40×2.44) unit. Although the magnitude of our estimate is small, it is encouraging that the NGOs' decision, on average, is not predominantly determined by the motivation to derive private benefits or to avoid disutility from exerting effort.

The proxy for altruism (RELIGIOUS) appears to have an insignificant association with diversification. Having their headquarters in Kampala is significantly associated with a lower number of activities performed, perhaps because most of these NGOs are local branches of international NGOs (63%) that already specialise in some pre-determined activities. Other control variables are consistent with expectations. Operating in more districts, having expanded geographically in the last five years, and having more staff are all correlated with more activities offered. There is also a U-shaped relationship between the manager's tenure and the number of activities. This correlation is consistent with the career concern hypothesis that new managers diversify to establish their reputation, but more established managers are reluctant to expand an organisation's portfolio because the benefit from signalling their ability to the job market is weaker (Gibbons & Murphy, 1992).

Table 4.4 presents estimates when we use the HHI index as the dependent variable. Columns 1 to Column 3 treat the HHI variable as continuous, assuming a linear relationship for our regressions. Under such an assumption, we obtain a similar pattern as reported in Table 4.3: once addressing omitted variable bias by including controls and using our flood instrumentation strategy, we observe a negative overall effect of incentives on the diversity index. In Column 4, we use an IV-Tobit estimator since the HHI is a limited dependent variable (censored at the lower end of its distribution) by construction, namely, the value of each HHI falls in the interval [0,1] and NGOs with no diversification are pooled with organisations with negative propensity to diversify at the lower end. The estimation provides asymptotically consistent and efficient estimates (Amemiya, 1984). Although we fail to replicate the significance as in Table 4.3 (probably due to the noises caused by a smaller sample size), the negative estimates imply that at least personal perks are not the dominant motivation underlying the diversification decision of NGOs in our sample. We focus our interpretation on the count index as its construction is more consistent to our theoretical setup.

Table 4.3. OLS and IV Estimations for diversification on incentives

VARIABLES	*Dependent variables:* Number of activities by classifications			
	(1)	(2)	(3)	(4)
INCENTIVES	0.301	0.10	-2.44**	-0.53**
	(0.235)	(0.23)	(1.07)	(0.22)
KAMPALA		-53.79***	-60.15***	-13.63**
		(19.62)	(22.35)	(5.33)
CHANGED FOCUS		37.40*	35.25	8.17
		(21.95)	(24.43)	(5.43)
DISTRICT		3.08	4.28	0.95*
		(2.50)	(2.66)	(0.52)
GEOGRAPHICAL EXPANSION		31.10	51.82**	11.34**
		(21.29)	(24.36)	(5.50)
LOGSTAFF		22.50**	26.75**	6.05**
		(9.64)	(11.15)	(2.49)
VOTE_ACTIVITY		-14.70	-35.40	-7.44
		(18.46)	(22.95)	(5.18)
RELIGIOUS TITLE		10.26	-10.25	-2.16
		(22.52)	(26.09)	(5.99)
TENURE		8.19**	10.06**	2.68**
		(3.50)	(4.14)	(1.07)
TENURE2		-0.26***	-0.29**	-0.08**
		(0.10)	(0.12)	(0.04)
OTHER_NGOS		-1.28***	-1.79*	-0.41*
		(0.48)	(1.06)	(0.24)
CONSTANT	410.316***	316.89***	454.32***	145.56***
	(16.721)	(35.85)	(70.72)	(13.48)
Estimator	OLS	OLS	2SLS	IV-Poisson
Anderson-Rubin confidence intervals		[-6.01, -0.51] p-value = (0.01)		
Observations	391	391	391	391

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Estimates are multiplied by 100 for ease of interpretation. The dependent variable is the number of activates in 2007. Two-step GMM estimator is used for the IV-Poisson. Anderson-Rubin confidence intervals are robust to weak instrumentation; the p-value is obtained after bootstrapping at 200 replications. See Table 4.1 for further explanations of control variables.

Table 4.4. OLS and IV Estimates for diversification on incentives

VARIABLES	Dependent variables: HHI index in 2007			
	(1)	(2)	(3)	(4)
INCENTIVES	-0.016	-0.02	-0.17	-0.19
	(0.031)	(0.03)	(0.15)	(0.16)
KAMPALA		-6.00**	-6.31**	-7.20**
		(2.76)	(2.90)	(3.11)
CHANGED FOCUS		1.49	1.71	1.97
		(2.89)	(2.86)	(3.21)
DISTRICT		2.26	0.43**	0.49
		(3.17)	(0.20)	(0.31)
GEOGRAPHICAL EXPANSION		0.35*	3.12	2.95
		(0.19)	(3.22)	(3.37)
LOGSTAFF		1.72	2.99**	3.16**
		(2.85)	(1.32)	(1.51)
VOTE_ACTIVITY		2.66**	0.80	0.95
		(1.27)	(2.88)	(3.01)
RELIGIOUS		1.75	-3.25	-3.68
		(2.66)	(3.56)	(3.75)
TENURE		-1.87	0.51	0.59
		(3.26)	(0.47)	(0.57)
TENURE2		0.44	-0.02	-0.02
		(0.44)	(0.01)	(0.02)
OTHER_NGOS		0.11	0.07	0.07
		(0.09)	(0.14)	(0.21)
CONSTANT	53.949***	44.50***	53.04***	52.69***
	(2.185)	(4.98)	(9.20)	(10.13)
Anderson-Rubin confidence intervals				[-0.63, 0.11] p-value = (0.00)
Estimator	OLS	OLS	2SLS	IV-Tobit
Observations	323	323	323	323

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Estimates are multiplied by 100 for ease of interpretation. Due to missing information on spending composition, only 323 NGOs are included when HHI index is used. the p-value is obtained after bootstrapping at 200 replications. See Table 4.1 for explanations of control variables.

4.4.4. Assessing the excludability assumption of the instrumentation

Our main identification assumption is that the location in the most affected areas only relates to the diversification decision through the increase in the proportion of contractual funding due to the surge from the international aid community and the decrease in the business incomes. Our assumption is that the funding surge and the flood would only affect the diversification decision only through the reliance on contractual funding. For one, it is unlikely that the NGOs in the affected areas could refocus their mission or introduce new activities regardless of whether they would receive funding from grants or memberships and user fees. The reason is that there would be no finance for such activities, given that the other income sources such as the founders' own funding or voluntary donations and business income would become limited following the depleted conditions in the areas. Only 9% of the managers were from a wealthy family; while on average 17% of these NGOs' total incomes came from non-contractual sources. For another, we also posit that the unexpected flood is exogenous to the characteristics of NGOs having a presence in affected areas prior to the event for two reasons. First, since the youngest NGO in our sample was founded before January 2007 (before the flood), no NGOs were founded simply to capture the new surge in funding. Second, in our specifications we control for whether the NGO expanded in the last five years to capture any opportunistic move into the affected areas in 2007. Three concerns remain.

First, as our diversification measures are recorded at the end of 2007, NGOs working in the affected areas may opportunistically aim to capture the new funding from the international aid community by introducing or shifting focus towards humanitarian and relief activities. If this is the case, our instrument excludability is violated since the change in the activity portfolio of these opportunistic NGOs is also directly affected by the flood, apart from the increased contractual funding from the international aid community (which is the IV mechanism). As such, having a presence in the most affected area could correlate with the number of activities recorded at the end of the year through parameters such as personal gains (λ) other than only through the increased INCENTIVE. We address this concern by two exercises. First, we control for whether the NGO has changed their focus in the last five years CHANGE_FOCUS. Second, we show that excluding NGOs working in humanitarian activities does not change our results. Indeed, among 104 NGOs reporting to have worked in the affected areas, only nine reported having activities in Shelter and Emergency Relief activities at the end of 2007. These activities present the best opportunities for organisations to capture the new sources of funding. Other activities, such as water and sanitation, health, education, employment and advocacy are long-term, locally rooted and plausibly not started as a response to the unexpected heavy rainfall. As

the NGOs with emergency activities could be the ones who opportunistically responded to the shock, we drop them from the main analysis and find the results similar in Table 4.A1.

The above argument does not invalidate FLOOD_AFFECTED being a strong instrument. According to the reports by UNOCHA (2008) and FAO (2008), humanitarian aid did not target only emergency missions but also other long-term activities. The first-stage mechanism would be that due to the flood NGOs in the most affected areas were able to attract funding to maintain and focus their current missions. The main argument against this mechanism is that NGOs in the most affected areas might be historically more reliant on revenue from charitable sources (incentives from the NGOs' stakeholders) regardless of the 2007 flood and the surge in funding. The argument means that the first-stage estimation merely reflects the location effect instead of the effect of the funding surge on the revenue composition, invalidating our IV mechanism. Section 4.4.5 uses a fixed effect model, controlling for whether the NGO expanded their geographical coverage from 2002 to 2007 and within-NGO variations, to address this concern. We show that even when we control for time-invariant characteristics (the location effect), variations in the reliance on charitable revenue (our INCENTIVES measure) over time remain negatively correlated with variations in the number of activities (the diversification measure).

The second concern is that the location decision of NGOs is rarely random: Some NGOs could self-select to operate in areas that are more prone to future extreme rainfalls and extreme flood events. For this argument to hold, one would need to show that: (i) NGOs could predict areas that are more likely to be affected by extreme rainfalls in the future; and (ii) there are systematic differences between NGOs who have worked in the most flood-prone areas and those who have not. We demonstrate that the two scenarios are unlikely to hold in our data.

First, there is no evidence that rainfall in Uganda is a predictable phenomenon, even between districts. Using monthly records for rainfall stations across Uganda from 1951 to 2003, Bjorkman-Nvqvist (2013) shows that she cannot reject the null hypothesis that rainfall in Ugandan districts follows a white-noise process. This finding suggests that there is no statistical correlation between the past amount of rainfall and the future. Consequently, this finding supports our argument that organisations are unlikely to locate in the flood-prone areas by observing indicators for such future incidents. The locations affected by the 2007 flood are unexpected and orthogonal to the characteristics of both the districts and the NGOs working in the affected areas.

Second, we demonstrate in Appendix 4 that there are mostly no systematic observable differences between organisations who have worked in areas prone to extreme floods and those who have not. We use the archive of global flood events obtained from the Dartmouth Flood Observatory (updated 2017) to measure the vulnerability to extreme

floods. That is, we define "an NGO working in areas vulnerable to extreme floods" as having a presence in districts identified by the Dartmouth archive as being hit by at least one extreme flood from 1998 to 2017. Excluding the 2007 event, the archive identifies 24 other extreme floods occurring in various districts of Uganda (see Table 4.A3 and Figure 4.A3 in Appendix). Using the Government of Ugandan's official district classifications in 2007, we identify 19 districts that have been affected by at least one extreme flood event during 1988 – 2017.[73] We code these districts as EVER_AFFECTED taking the value of 1 if the district has been hit by at least one extreme flood during 1988 – 2017, 0 otherwise. Our balance tests in Appendix 4.6.3 works as follows. If some NGOs manage to self-select into areas most vulnerable to a future flood, we must also observe the selection between NGOs working in areas that are historically affected by extreme floods: areas experiencing extreme floods before 2007 would indicate the higher risk of being hit again in the future. Besides, it is plausible to expect areas that were affected by floods after 2007 would equally attract the NGOs who selectively located in these areas before the 2007 flood. The rationale behind this argument is that although NGOs could selectively locate in areas before 2007 in the hope that these areas would be affected by an extreme flood that would require a sizable humanitarian support from the international aid community *at one time* in the future, it is unlikely that they could pin down when. Figure 4.3 supports this argument – the rainfall level in 2007 was at least twice as much the second largest rainfall level ever recorded in Uganda. Assuming the timing (the year 2007) of the flood is unexpected, we compare NGOs who have a presence or have not a presence in areas most vulnerable to extreme floods between 1988 and 2017. If there is no observable difference concerning the NGOs characteristics between NGOs who have and have not worked at the vulnerable areas, we could rule out the possibility that some NGOs self-predict and self-select into flood-prone areas. The balance also alleviates the concern that international donors may target some NGOs, who concurrently have worked in the affected areas in 2007. In that case, our instrument may be correlated with characteristics of the NGO apart from the working location. Appendix 4.6.3 shows that it is not a practical concern.

The third threat is that NGOs working in affected areas may differ from those in unaffected areas in unobservable ways. There may be some unobservable characteristics of the districts most vulnerable to floods and our specifications fail to control for associated selectivity. To address this concern, we undertake a robustness check in which we restrict

[73] In 2010, the Act of Local Government was enacted. Several previous districts were merged, and others were separated to form new official districts. We conservatively restrict our analysis to the administrative boundaries in 2008 when our survey took place. For districts identified in the period 2010-2012, we match them with the pre-2010 official districts.

our sample to organisations working in areas that are historically vulnerable to floods (EVER_AFFECTED = 1). By doing so, we compare NGOs operating in areas that are equally likely to be affected by an extreme flood, where we let the timing and scale of the 2007 flood be exogenous and unexpected. Consequently, absent of prior knowledge of the 2007 shock, these NGOs would have similar expectations regarding the uncertainty of the environmental effect and private benefits following any future increase in international aid relief. Since the timing is plausibly unpredictable, the 2007 flood instrument causes exogenous variations in the grant package due to the increased difficulty of working in the severely affected areas. Any difference in activity diversification in 2007 of these NGOs (with EVER_AFFECTED = 1) would be caused by these exogenous variations in their reliance on contractual grants, not due to prior intention to locate at flood-prone areas.

Table 4.5 presents estimates for the restricted sample of 280 NGOs who are identified as operating in areas historically affected by at least one flood during 1988 - 2017. We observe a similar negative and significant effect of incentives on the diversification level for our preferred specification using the count measure. Estimates of other control variables are also consistent with that in the full sample. The results demonstrate that our observed negative effect is not driven by self-selection into locations that may experience extreme weather and therefore be more likely to attract funding. [74]

[74] One related concern is that due to the cease-fire treaty between the Ugandan Government and the Lord's Resistance Army in 2006-2008, our first-stage estimation might pick up the increase in the aid or the government reconstruction support towards districts where the conflicts were settled. Given the survey design, it is not an issue. The initial survey took place in 2002 when it was unsafe for enumerators to travel to the affected areas and hence no interviews were conducted with NGOs working in the conflict areas (mainly in North Uganda). Furthermore, when we exclude the two NGOs that report construction-related activities, the results remain.

Table 4.5 IV estimates using restricted sample: EVER_AFFECTED = 1

VARIABLES	Count measure		HHI index	
	(1)	(2)	(3)	(4)
INCENTIVES	-2.14**	-0.47**	-0.20	-0.22
	(1.08)	(0.23)	(0.16)	(0.18)
KAMPALA	-46.49*	-10.43*	-5.53*	-6.45*
	(23.99)	(5.75)	(3.33)	(3.79)
CHANGED FOCUS	35.94	8.35	3.07	3.30
	(27.98)	(6.31)	(3.62)	(4.09)
DISTRICT	4.58*	1.02*	0.55**	0.62*
	(2.70)	(0.52)	(0.22)	(0.34)
GEOGRAPHICAL EXPANSION	37.56	8.32	-0.75	-1.24
	(28.03)	(6.52)	(3.92)	(4.19)
LOGSTAFF	29.65**	6.93**	4.15**	4.47**
	(12.90)	(2.98)	(1.68)	(1.91)
VOTE_ACTIVITY	-37.20	-7.91	1.04	1.19
	(28.73)	(6.58)	(3.97)	(4.10)
RELIGIOUS	20.66	5.06	-3.43	-4.10
	(29.30)	(6.75)	(4.19)	(4.61)
TENURE	8.80*	2.28**	1.00*	1.10
	(4.56)	(1.13)	(0.54)	(0.68)
TENURE2	-0.23**	-0.06**	-0.03**	-0.03
	(0.11)	(0.03)	(0.01)	(0.02)
OTHER_NGOS	1.54	0.35	0.69**	0.75
	(1.87)	(0.45)	(0.35)	(0.53)
CONSTANT	415.53***	137.25***	48.42***	47.72***
	(76.52)	(15.59)	(10.51)	(11.79)
Estimator	2SLS	IV-Poisson	2SLS	IV-Tobit
Observations	280	280	229	229

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Estimates are multiplied by 100 for ease of interpretation. Two-step GMM is used for IV-Poisson. See Table 4.1 for explanations of control variables.

4.4.5. Alternative identification strategy

In the main analysis, we identify the incentive effect using between-NGO variations in the reliance on income from stakeholders after the 2007 flood. A different approach would be to use variations within organisations (fixed effects) to control for the unobservable parameters that jointly determine the level of diversification and the incentive measures. The fixed effects also plausibly capture the systematic misreporting (recalling) of the NGO representative and the location effect discussed in the previous section.

To construct a fixed-effect estimation, we exploit several 2008 survey questions that ask the respondent to recall: (i) activities introduced since 2002 ($n_{i2007_{introduced}}$), and

(ii) activities discontinued since 2002 ($n_{i2007_{discontinued}}$). We exploit within-organisation variations in the level of diversification over two periods in an attempt to identify the average incentive effect of interest. The two-way fixed effect model is as follows:

$$n^*_{it} = \beta \text{INCENTIVES}_{\text{it}} + \text{X}'_{\text{it}}\alpha + \pi_{\text{i}} + \delta_t + \varepsilon_{\text{it}} \tag{18}$$

where $t = 2002, 2007$. $n_{i2002} = n_{i2007} + n_{i2007_{introduced}} + n_{i2007_{discontinued}}$. As the survey specifically asks for revenue information in both 2002 and 2007, the construction of $\text{INCENTIVES}_{\text{i2002}}$ is like that of $\text{INCENTIVES}_{\text{i2007}}$. For those NGOs who were not around in 2002 or the respondents could not recall the information (22 NGOs), we code n_{i2002} and $\text{INCENTIVES}_{\text{i2002}}$ as missing.

Dummies π_{i}, δ_t are for organisation and time fixed effects. π_i captures time-invariant firm specific characteristics of the NGO, which are likely to capture the unobserved exogenous parameters: risk aversion (r), private benefit (λ), cost of diversification (β), and disutility aversion (k) of the organisation leadership due to working location. The fixed effect also captures the concern of systematic reporting bias since the information was recalled by the same individual. Second, δ_t captures the external yearly characteristics that may affect both the $\text{INCENTIVES}_{\text{i}}$ and n, for example, government policies toward the non-profit sector or nation-wide economic shocks. As such, δ_t also captures the external volatility parameter (s^2) faced by the NGO specified in the theory.

X_{it} is a set of time-variant variables that may affect the four parameters at the organisational level. The first source is that the characteristics of the leadership r, λ, β, k could change over time due to not only nation-wide time trends but changes in the leadership itself. We control for this source by including a binary variable NEW_MANAGER2007 indicating whether the manager in 2007 has been with the NGO since 2002. The variable takes the value of 1 at $t = 2007$ and 0 at $t = 2002$ if the manager was new and taking the value of 0 at both $t = 2007$ and $t = 2002$ if the manager has been with the NGO since 2002. Private benefits and disutility of project choice could change over time even under the same management if the NGO underwent any change in their focus or geographical coverage. We control for these possibilities by including two binary variables (i) CHANGED FOCUS takes the value of 1 at $t = 2007$ if the NGO has changed its focus since 2002, 0 otherwise; and (ii) GEOGRAPHICAL EXPANSION takes the value 1 at $t = 2007$ if the NGO has expanded its geographical coverage since 2002, 0 otherwise. To account for the change in the size of the NGOs, we use the number of staff in 2002 and 2007 (NUMSTAFF) as the measures. The construction of the number of staff in 2002 is similar to that of $\text{INCENTIVES}_{\text{i2007}}$. Due to missing information, we end up with an unbalanced panel of the two periods with 369 NGOs and 517 NGO-year units of

observation. To avoid introducing appreciable bias to our sample, we do not drop singleton observations in the main analysis.

Table 4.6 reports similar results as our previous estimates. Overall, after controlling for the time and organisation fixed effects, there still exists a significant and negative effect of the revenue from contractual funding on the decision to diversify. The effect remains when we use a panel IV-Poisson estimation to capture the count nature of the dependent variable in Column 5. The robustness of the results supports our estimates using between-NGO variations. We also observe a strong time fixed effect over the period, suggesting that Ugandan NGOs tended towards a higher level of diversification from 2002 to 2007. These results support the need to control for unobserved characteristics that jointly determine the level of diversification and the percentage of revenues from contractual sources suggested in our theoretical model. Since we do not have information on expenditure on activities in 2002, we are unable to use the HHI as an alternative dependent variable.

Table 4.6. Panel Estimations of diversification on incentives

VARIABLES	Dependent variables: Number of activities by classifications				
	(1)	(2)	(3)	(4)	(5)
INCENTIVES	0.006	-0.04	-0.25*	-0.36**	-0.09**
	[0.156]	[0.15]	[0.14]	[0.18]	[0.04]
TREND (2007 = 1)		31.68***	41.68***	77.22***	16.70***
		[10.24]	[10.56]	[24.63]	[5.62]
GEOGRAPHICAL EXPANSION				-8.70	-1.27
				[22.86]	[5.08]
CHANGED FOCUS (Yes = 1)				-62.21***	-13.67***
				[20.22]	[4.45]
NEW_MANAGER2007 (Yes = 1)				8.05	1.84
				[28.20]	[6.15]
NUMSTAFF				0.17	0.04
				[1.39]	[0.33]
CONSTANT	422.242***	400.10***	414.95***	417.25***	
	[14.063]	[16.11]	[11.79]	[38.18]	
Time FE	-	Yes	Yes	Yes	Yes
Organisation FE	-	-	Yes	Yes	Yes
OLS/Poisson	OLS	OLS	OLS	OLS	Poisson
Observations	517	517	517	517	296
Number of groups	369	369	369	369	148

Notes: *** p<0.01, ** p<0.05, * p<0.1. Unit of observations is NGO-year. Robust standard errors are in brackets. Estimates are multiplied with 100 for ease of interpretation. The number of activities is reported by each organisation that they operated in 2002 and 2007. The activities are classified by categories to avoid overlap. INCENTIVES is measured as revenue from Local grants, International grants, Membership and User fees as a percentage of total income. NEW_MANAGER2007 takes the value 1 at $t = 2007$ if the manager was new in 2007, 0 otherwise; CHANGED FOCUS takes the value 1 at $t = 2007$ if the NGO has changed its focus since 2002, 0 otherwise; GEOGRAPHICAL EXPANSION takes the value 1 at $t = 2007$ if the NGO has expanded its geographical coverage since 2002, 0 otherwise. NUMSTAFF measures the number of staff in 2002 and 2007. We have fewer observations for the Poisson fixed-effect estimator because NGOs who did not experience a change in the diversification level are automatically dropped from the estimation.

4.5. Concluding remarks

NGOs play an increasingly important role in aid disbursement and project implementation around the world. As such, understanding the motivation for diversifying the activities NGOs undertake is important when devising development projects. We identify two motivations for diversification in the activities that NGOs in Uganda undertake. We incorporate these motivations, namely to reduce uncertainty of financial flows and to gain

private benefits, in a theoretical framework and derive predictions regarding how an NGO responds to a change in incentives offered by their stakeholders. We predict that a higher incentive, namely increased contractual funding, will decrease diversification for NGOs motivated by reducing the uncertainty; while a higher incentive will increase diversification for NGOs motivated by private gain. Exploiting variations in the funding after the surge of international grants toward several Ugandan regions following an unexpected flood in 2007, we show that NGOs respond to an increase in the proportion of incomes from contracts (such as grants, membership or user fees) by lowering the number of activities they offer. The result is robust when we control for within-organisation variations, time fixed effects, and several time-varying characteristics using additional information of the NGO's activities and funding in 2002.

According to our model, the empirical evidence is consistent with the interpretation that Ugandan NGOs, on average, diversify primarily to reduce risks associated with their operation rather than to derive personal gains. This implies NGOs themselves consider diversification not to be a desirable action in its own right. As such, policies should be directed to ensure NGOs' financial resilience, particularly for grassroots NGOs who mainly rely on external funding. If the underlying reason for diversification is to improve financial stability, stakeholders can help NGOs focus on their overriding mission by providing a commitment to funding. Our result is encouraging as we do not find supporting evidence for the hypothesis that the motivation underlying NGOs' behaviour is self-benefitting, at least concerning activity diversification. As such, donors could design their aid and grant package conditions that directly target the financial stability of their recipients and subcontractors as a way to improve the efficiency of the NGO sector.

4.6. Appendix For Chapter 4

4.6.1. Derivation of the Certainty Equivalent (Equation 4)

Given $\varpi = w_0 + \alpha[e + \varepsilon(n)] + \lambda \ln n - \beta n - \frac{ke^2}{2}$ and $\epsilon(n)$~N$(0, \frac{\sigma^2}{2})$ the induced wealth is also normal distributed with mean of $\mu = w_0 + \alpha e + \lambda \ln n - \beta n - \frac{ke^2}{2}$ and variance of $s = \frac{\alpha^2\sigma^2}{n}$. The corresponding density function is:

$$f(\varpi) = \frac{1}{s\sqrt{2\pi}} \exp(-\frac{(\varpi - \mu)^2}{2s^2})$$

Since $u(\varpi) = -\exp(-r\varpi)$, the corresponding expected exponential utility function is:

$$
\begin{aligned}
E(u(\varpi) &= -\int_{-\infty}^{+\infty} \exp(-r\varpi) f(\varpi) d\varpi \\
&= -\int_{-\infty}^{+\infty} \exp(-r\varpi) \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(\varpi-\mu)^2}{2s^2}\right) d\varpi \\
&= -\frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-r\varpi - \frac{(\varpi-\mu)^2}{2s^2}\right) d\varpi
\end{aligned}
$$

The inside function is:

$$
-r\varpi - \frac{(\varpi-\mu)^2}{2s^2} = -\frac{1}{2s^2}[(\varpi-\mu) + rs^2]^2 - r\mu + \frac{rs^2}{2}
$$

Substitute that to $E(u(\varpi)$ we have:

$$
\begin{aligned}
E(u(\varpi) &= -\frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2s^2}[(\varpi-\mu) + rs^2]^2 - r\mu + \frac{rs^2}{2}\right) d\varpi \\
&= -\exp(-r\mu + \frac{rs^2}{2}) \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2s^2}[(\varpi-\mu) + rs^2]^2\right) d\varpi
\end{aligned}
$$

Let

$$
\begin{aligned}
g(\varpi) &= \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2s^2}[(\varpi-\mu) + rs^2]^2\right) d\varpi \\
&= \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2s^2}[\varpi - (\mu - rs^2)]^2\right) d\varpi
\end{aligned}
$$

be the density function of a random variable that follows a normal distribution with mean of $\mu - rs^2$ and variance of s^2. Therefore:

$$
\frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2s^2}[\varpi - (\mu - rs^2)]^2\right) d\varpi = 1
$$

$$
E\big(u(\varpi)\big) = -\exp\left(-r\mu + \frac{r^2 s^2}{2}\right) = -\exp\left[-r\left(\mu - \frac{rs^2}{2}\right)\right] = -\exp(rCE)
$$

Hence the certainty equivalent is

$$
CE = w_0 + \alpha e + \lambda \ln n - \beta n - \tfrac{ke^2}{2} - \tfrac{rs^2}{2}
$$

4.6.2. Proof of Corollary 2

The proof slightly follows the approach in Aggrawal and Samwick (2003).

We prove that when k is sufficiently high, we have:

$$
\frac{\partial \alpha^*}{\partial j} < 0, j \in \{r, \sigma^2, k, \beta\} \qquad \text{and} \qquad \frac{\partial \alpha^*}{\partial \lambda} > 0
$$

$$
\frac{\partial n^*}{\partial j} > 0, j \in \{r, \sigma^2, \lambda\} \qquad \text{and} \qquad \frac{\partial n^*}{\partial k}, \frac{\partial n^*}{\partial \beta} < 0
$$

First, recall that α^* is the unique solution to:

$$
0 = \frac{1-\alpha}{k} + f(k) - \frac{2\alpha\beta r\sigma^2}{\sqrt{\lambda^2 + 2\beta r\sigma^2\alpha^2} + \lambda} \qquad (A1)
$$

We first prove the statics for α^*. Simplify (A1) we have:

$$t(\alpha) = \frac{2\alpha\beta r\sigma^2}{\sqrt{\lambda^2 + 2\beta r\sigma^2\alpha^2} + \lambda} = \frac{1-\alpha}{k} + f(k) \tag{A2}$$

Since λ appears only in the denominator of (A2), $t(\alpha)$ is increasing in λ.

Rewrite $t(\alpha)$ as:

$$t(\alpha) = \frac{2\beta r\sigma^2}{\sqrt{\frac{\lambda^2}{\alpha^2} + 2\beta r\sigma^2} + \frac{\lambda}{\alpha}} \tag{A3}$$

It is trivial now to see $t'(\alpha) > 0$. To derive ${\partial\alpha^*}/{\partial k}$, we use the Implicit function theorem for (A2). Assuming that k is sufficiently high (or the NGOs work in a challenging environment) and $f'(k) > 0$, we will have $f'(k)k^2 > 1 - \alpha$. Since $\alpha < 1$ and $t'(\alpha) > 0$, we have the following:

$$\frac{\partial\alpha}{\partial k} = -\frac{f'(k) + \frac{\alpha - 1}{k^2}}{-\frac{1}{k} - t'(\alpha)} > 0$$

Now, rearrange (A1) as:

$$\left(\frac{1-\alpha}{k} + f(k)\right)\frac{\sqrt{\lambda^2 + 2\beta r\sigma^2\alpha^2} + \lambda}{\alpha} = 2\beta r\sigma^2$$

Intuitively, when holding k, r, σ^2, β constant, raising λ must be compensated by raising, though not proportionately, α. Hence, intuitively $\frac{\partial\alpha}{\partial\lambda} > 0$. Again, when holding k, λ constant, raising any component of (β, r, σ^2) would require a decrease in α to ensure the equation. Thus, $\frac{\partial\alpha}{\partial\beta r\sigma^2} > 0$.

For the statics of n^*, recall the formula for the optimal diversification as:

$$n^* = \frac{\lambda}{2\beta} + \frac{1}{2}\sqrt{\left(\frac{\lambda}{\beta}\right)^2 + 2\frac{r}{\beta}s^2[\alpha^*(r, s^2, \lambda, k, \beta)^2]} \tag{A4}$$

Holding others constant, we have $\frac{\partial n^*}{\partial\lambda} > 0$ and $\frac{\partial n^*}{\partial\beta} < 0$. Since $\frac{\partial n^*}{\partial\alpha} > 0$ and $\frac{\partial\alpha^*}{\partial k} > 0$, we have $\frac{\partial n^*}{\partial k} > 0$.

Let $T = r\sigma^2$, to show $\frac{\partial n^*}{\partial j} > 0, j \in \{r, \sigma^2\}$, it suffices to show that $\frac{\partial n^*}{\partial T} > 0$. Indeed, differentiate (A4) with respect to T:

$$\frac{\partial n^*}{\partial T} = \frac{1}{\beta}\alpha^2 \times$$

$$\times \frac{2T^2\alpha^4\beta^2 - 4T^2\alpha^2\lambda\beta + 5T\alpha^2k\lambda\beta^3 + 3T\phi^3k\lambda^2 - 2\beta\phi k\lambda^3\beta^2 + 2\lambda^4\beta + 2\phi\lambda^3k\beta^2 - 2T^2\alpha^2\phi^2\beta^2}{\phi(2T^2\alpha^4\beta + 2T^2\alpha^2k\lambda + 5\lambda^2T\alpha^2\beta^2 + 3T\alpha^2\phi\lambda + 2T\phi k\lambda + 2Tk\lambda^3\beta^2 + 2\lambda^4\beta^2 + 2\phi\lambda^3\beta}$$

where $\phi = \sqrt{(\lambda^2\beta^2 + 2T\alpha^2\beta^4)}$.

The denominator is positive as it is a sum of positive expressions, we can show that the nominator is positive too if we assume k is sufficiently high and $\phi > \lambda\beta + \alpha\beta^2$.

Hence ${\partial n^*}/{\partial T} > 0$. The proof completes.

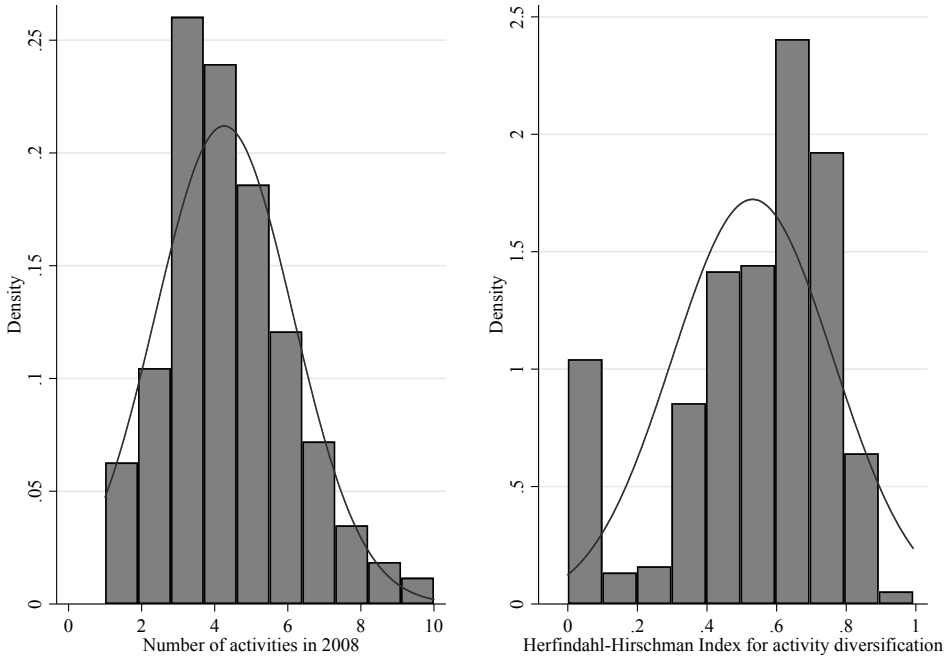
4.6.3. Additional Tables and Figures

Table 4.A1. Robustness to excluding NGOs working on Shelter and Emergency Relief

Variables	*Dependent variables*					
	Count index			HHI		
	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Incentives	0.08	-2.42**	-2.11**	-0.02	-0.21	-0.23
	(0.23)	(1.01)	(1.03)	(0.03)	(0.16)	(0.17)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Restricted	Full	Full	Restricted

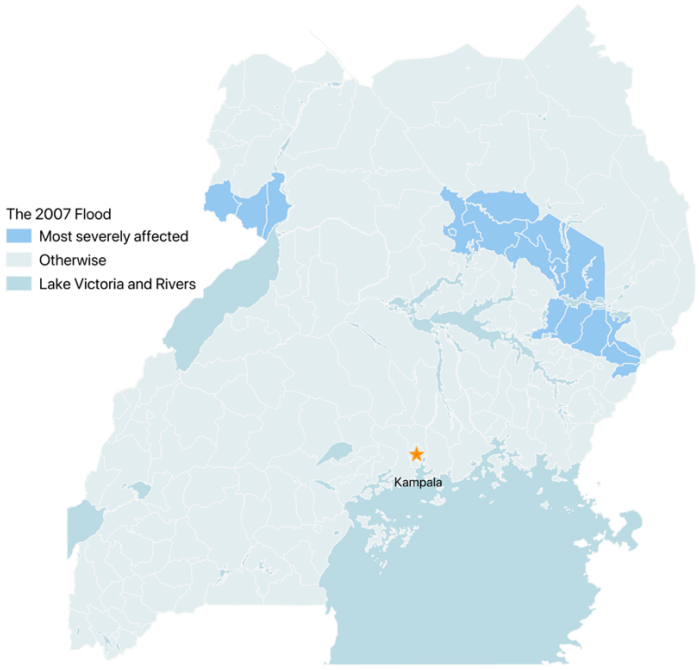
Note. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Estimates are multiplied by 100 for ease of interpretation. The samples used are either full (including all available NGOs) or restricted (including NGOs most vulnerable to an extreme flood during the period 1988 – 2017). All estimations are with the list of controls specified in Table 4.3. To demonstrate the robustness to the case when NGOs opened new activities related to Shelter and Emergency Relief as a response to the flood in 2007, this table reports results when we further exclude 9 NGOs who reported participation in the activity from the sample.

Figure 4.A1. Histograms of the two dependent variables



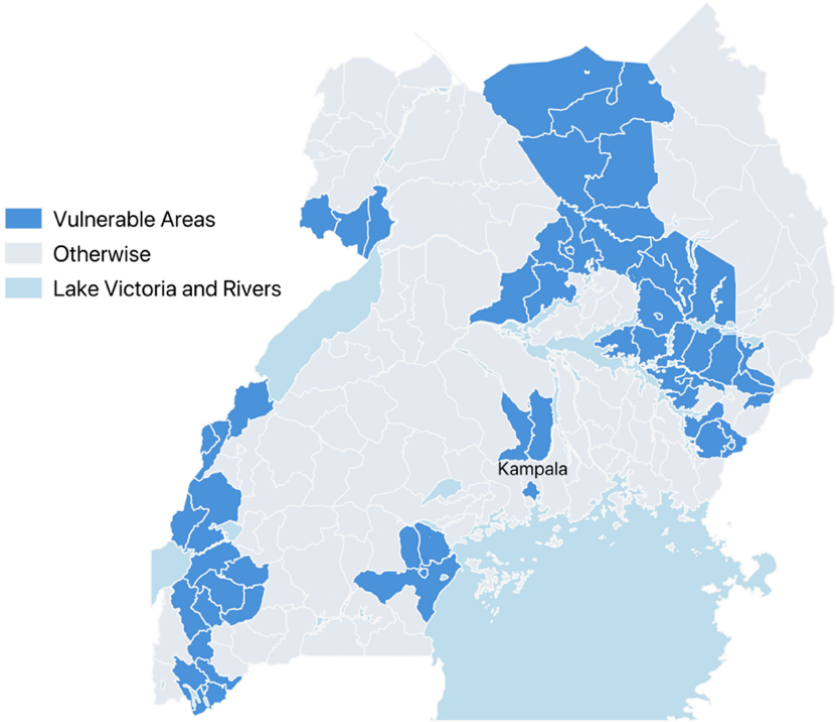
Note. The number of activities and the Herfindahl-Hirschman index are constructed using the 2008 Ugandan NGO survey. Bars represent the variable's density. Lines represent the density of normal distributions.

Figure 4.A2. Ugandan districts most affected by the 2007 flood



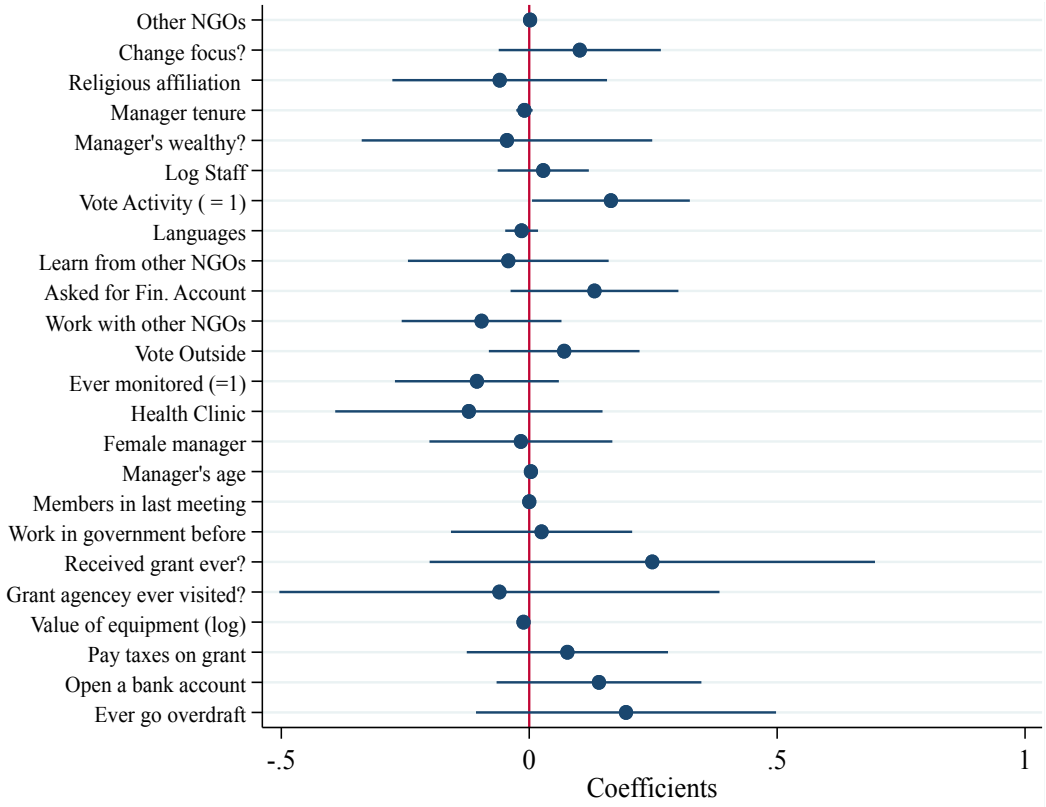
Note. The most affected areas are identified by the UNOCHA and the Ugandan Red Cross. The official list includes Soroti, Amuria, Katakwi, Bukedea, Kumi, Lira, Sironko, Bududa and Nebbi. The geographical boundaries are as of 2006.

Figure 4.A3. Ugandan districts most vulnerable to extreme floods from 1988 to 2017



Note. Districts vulnerable to extreme floods are defined as places that have been hit by at least one extreme flood during from 1988 to 2017 as recorded in Table 4.OA1. The data are obtained from the Dartmouth Archive of Global Flood Events (updated 2017).

Figure 4.A4. OLS regression of FLOOD_AFFECTED on NGOs characteristics



Note. We report an OLS regression of FLOOD_AFFECTED over a range of observable NGO characteristics. No particular characteristics have significant power to predict whether the NGO would work in a district severely hit by the 2007 flood. The result reinforces our argument that it is unlikely for Ugandan NGOs to self-select to flood-prone areas. For details of the variables, see Table 4.A4. FLOOD_AFFECTED = 1 if the NGO had the presence or worked at a district hit by the 2007 flood, 0 otherwise.

4.6.4. Balance tests and additional tables

Table 4.A2 provides evidence of balance between NGOs that have and have not worked in vulnerable areas, defined as districts hit by at least one extreme flood from 1988 to 2017 recorded in the Dartmouth Flood Observatory. The result shows no significant between the NGOs' characteristics. Figure 4.A4 above further reinforces the idea by showing that no NGO's characteristics have any significant power to predict the location of the NGO at affected areas by the 2007 flood.

Table 4.A2. Balance checks on working in vulnerable areas

	Not worked	Worked	Difference	t statistics	p-values
Other NGOs	1.85	1.42	0.43	0.57	0.57
Change focus?	0.30	0.21	0.08	1.95**	0.05
Religious affiliation	0.21	0.24	-0.03	-0.64	0.52
Manager tenure	7.47	6.52	0.95	1.67	0.10
Manager's wealthy?	0.10	0.07	0.03	1.01	0.31
Languages	3.78	3.40	0.38	1.15	0.25
Log Staff	2.62	2.67	-0.05	-1.24	0.14
Learn from other NGOs	0.12	0.15	-0.03	-0.80	0.42
Asked for Fin. Account	0.60	0.54	0.06	1.22	0.22
Work with other NGOs	0.49	0.51	-0.02	-0.35	0.73
Vote Activity (=1)	0.34	0.36	-0.02	-0.40	0.69
Ever monitored (= 1)	0.66	0.62	0.04	0.73	0.46
Health clinic (=1)	0.15	0.12	0.02	0.68	0.49
Female manager	0.22	0.27	-0.05	-1.07	0.28
Manager's age	43.46	43.02	0.44	0.39	0.70
Members in last meeting	142.95	116.55	26.40	0.53	0.60
Work in government before	0.48	0.40	0.08	0.05	0.13
Received grant ever?	0.65	0.64	0.01	0.20	0.84
Grant agency ever visited?	1.07	1.06	0.01	0.27	0.78
Value of equipment (log)	5.88	6.03	-0.15	-0.22	0.82
Pay taxes on grant?	0.15	0.11	0.05	1.38	0.17
Open a bank account?	0.70	0.77	-0.07	-1.50	0.13
Ever go overdraft?	0.05	0.09	-0.04	-1.28	0.20
Number of NGOs	198	280			

Notes: For completeness, we report balance checks for the whole sample of 478 NGOs. T-tests of characteristics between NGOs working in areas identified as ever affected by at least one extreme flood as in the Archive of Global Flood Events (Dartmouth Flood Observatory) during 1988-2017. Variables' names are self-explaining.

Table 4.A3. Extreme floods across Ugandan districts from early 1980 to 2017

ID	Detailed Locations	Began	Ended	Dead	Displaced	Main cause	Severity	Affected km^2
1	Acholi districts, Agago, Kitgum, Lamwo and Pader, northern Uganda	20-Aug-12	7-Sep-12	0	15000	Heavy Rain	1	57981.45
2	Butaleja district	20-May-12	29-May-12	0	3000	Heavy Rain	1	43315.84
3	Sere District	9-Nov-11	20-Dec-11	0	15000	Torrential Rain	1.5	22324.63
4	Bulambuli District in eastern Uganda	20-Aug-11	8-Sep-11	31	0	Heavy Rain	1.5	37846.35
5	Mazimansa, Himutu and Kachonga sub-counties	26-Sep-10	30-Sep-10	0	600	Heavy Rain	1.5	3303.97
6	Mutumba Zone, Kigwo and Kimanyika villages	7-Mar-10	12-Mar-10	0	20000	Heavy Rain	1	75196.57
7	Kampala city - Lufuka, Kajjansi and Abayita Ababiri in Wakiso district	15-Nov-07	18-Nov-07	4	3000	Heavy rain	1	552.247
8	Soroti, Amuria, Katakwi, Bukedea, Kumi, Lira, Sironko, Bududa (Mbale) and Nebbi	15-Aug-07	31-Oct-07	52	520000	Heavy rain	1.5	107797.14
9	Sironko District - Bumufuni, Bunambutye: Busangai, Bukhayaki, Bumbocha, Bumuyonga, Kapchorwa	12-Aug-06	21-Aug-06	3	1680	Heavy rain	1	4902.34
10	Pakwach area in Nebbi district. Apac town	8-Nov-04	10-Nov-04	0	10000	Heavy rain	1	299.295
11	Kasese District - Karusandara sub-county - Kanamba parish: (Mubuku II, Ngando and Kanyatete). Ibuga parish: Kikoga and Nyabubale	18-Oct-04	20-Oct-04	0	1500	Heavy rain	1	1127.99
12	Eastern Uganda - Mbale area - Bubulo county - Buwabwala sub-county, Busiu and Bukibino	1-Jul-03	3-Jul-03	20	700	Heavy rain	1	962.615
13	Uganda - Kampala area, Nakivubo Channel. Eastern districts: Mbale, Bugiri, Sironko. Southwestern districts: Rukungiri, Kabale and Busheny. Kenya - Nyando District, Ayweyo, Awach River.	16-Nov-02	25-Nov-02	2	2300	Heavy rain	1	15002.61
14	Masaka district: Kyazanga	15-Nov-01	16-Nov-01	0	1200	Heavy rain	1	1218.62
15	Villages around Rwebisengo headquarters. Bundibugyo town. River Semliki	20-Aug-01	29-Aug-01	0	4000	Heavy rain	1	15804.14
16	Apac district - Lake Kwania	20-Jul-98	24-Jul-98	0	20000	Heavy rain	1	3408.41
17	Mbale district, Tororo district	14-Nov-97	28-Nov-97	35	20000	Heavy rain	1	20737.67
18	Kasese district - Kyarumbar subcounty, Rwenzori Mountain slopes	30-Apr-94	2-May-94	11	0	Heavy rain	1	9570.23
19	Bundibugyo district (about 200 miles west of Kampala)	23-Nov-88	7-Dec-88	0	5000	Torrential rain	1	14344.36
20	"various parts of the northern region"	11-Aug-88	25-Aug-88	5	0	Torrential rain	1	55984.41

Notes: Data obtained from the Archive of Global Flood Events, the Dartmouth Flood Observatory (updated 2017) for Uganda. There are three classes of severity. Class 1 includes large flood events with significant damages to structures or agriculture. Class 1.5 includes very large events: with a greater than 2 decades but less than 100-year estimated recurrence interval, and/or a local recurrence interval of at 1-2 decades and affecting a large geographic region ($>$ 5000 sq. km). Class 2 includes extreme events with an estimated recurrence interval greater than 100 years.

Table 4.A4. The number of NGOs working in each district in 2007

District	NGOs	District	NGOs	District	NGO	District	NGOs	District	NGOs
Kampala	211	Amuru	49	Mpigi	27	Manafwa	20	Dokolo	15
Abim	182	Amolatar	44	Nakasongola	27	Moyo	20	Kanungu	15
Bududa*	94	Gulu	44	Kabarole	26	Mubende	20	Kalangala	14
Arua	86	Mayuge	43	Nakapiripirit	26	Ntungamo	19	Kaberamaido	13
Mukono	83	Bugiri	42	Oyam	26	Ibanda	18	Kapchorwa	13
Apac	76	Masaka	42	Pader	26	Kiboga	18	Kotido	13
Jinja	76	Sironko*	41	Kayunga	25	Rakai	18	Lyantonde	12
Kasese	74	Busia	40	Kitgum	25	Isingiro	17	Moroto	12
Mbale	71	Adjumani	39	Sembabule	24	Koboko	17	Rukungiri	12
Iganga	68	Kamuli	37	Bundibugyo	23	Namutumba	17	Butaleja	11
Tororo	65	Pallisa	37	Kibaale	22	Kamwenge	16	Kaabong	11
Budaka	61	Kabale	35	Kisoro	22	Kiruhura	16	Mityana	11
Lira*	60	Bushenyi	32	Nakaseke	22	Kyenjojo	16	Bukedea*	9
Amuria*	53	Masindi	29	Katakwi*	21	Maracha-Terego	16	Bulisa	8
Mbarara	53	Soroti*	29	Hoima	20	Nebbi*	16	Bukwa	6
Luweero	52	Kumi*	27	Kaliro	20	Wakiso	16	Yumbe	5

Note. The number of NGOs working in each district at the end of 2007. Data were drawn from the 2008 Ugandan NGO survey by Burger and Owens (2011). We sort the table in a descending order. Figure 4.2 presents a graphical depiction of this table. * represents districts severely affected by the 2007 flood depicted in Figure 4.A2.

Table 4.A5. Balance checks for NGOs on revenue composition

	Without	With	Difference	t statistics	p-values
No. of activities in 2007	4.13	4.29	-0.16	-0.72	0.47
No. of new act. in 2002	0.53	0.73	-0.20	-1.72	0.09
No. of discontinued act. in 2002	0.40	0.36	0.03	0.33	0.74
Other NGOs	1.98	1.44	0.53	0.61	0.54
Kampala Headquarter (=1)	0.44	0.39	0.05	0.84	0.40
No. of working districts	4.41	3.94	0.47	0.60	0.55
Change focus (=1)	0.20	0.24	-0.04	-0.88	0.38
Number of staff (log)	2.78	2.80	-0.02	-0.14	0.89
Religious affiliation	0.26	0.23	0.03	0.57	0.57
Manager Tenure	7.39	6.67	0.72	1.04	0.30
Manager's wealthy	0.06	0.08	-0.02	-0.65	0.51
Taxes on Grant	0.14	0.12	0.02	0.47	0.64
Having a bank account (=1)	0.81	0.73	0.08	1.52	0.13
Having overdraft	0.10	0.08	0.02	0.70	0.48
Languages	3.26	3.78	-0.52	-1.46	0.15
Learn from other NGOs	0.20	0.13	0.06	1.51	0.13
Evaluated by other NGOs	0.57	0.58	-0.01	-0.13	0.90
Asked for Fin. account	0.52	0.56	-0.04	-0.72	0.47
Worked with other NGOs	0.44	0.52	-0.07	-1.21	0.23
Has NGOs as member (= 1)	0.17	0.24	-0.07	-1.30	0.19
Vote Activity (=1)	0.26	0.37	-0.11	-1.94	0.05
Monitored (= 1)	0.64	0.63	0.01	0.23	0.82
International Travel (= 1)	0.64	0.63	0.01	0.23	0.82
Manager has uni. degree (=1)	0.83	0.86	-0.03	-0.81	0.42
Value of equipment (log)	15.72	15.64	0.07	0.25	0.80
Value of rented vehicle (log)	5.48	6.10	-0.62	-0.77	0.44
Belong to a global NGO	0.10	0.07	0.03	1.07	0.28
Health Clinic (=1)	0.10	0.14	-0.04	-0.89	0.37
Female manager (= 1)	0.22	0.27	-0.04	-0.81	0.42
No. of members in last meeting	0.10	0.14	-0.04	-0.89	0.37
Find needs of community	0.22	0.27	-0.04	-0.81	0.42
Ever received grant (= 1)	0.68	0.63	0.04	0.68	0.49
Number of NGOs	87	391			

Note. T-tests of characteristics between 87 NGOs without information revenue composition (excluded from the analysis) and 391 NGOs with the information. We demonstrate that the missing information is likely due to random factors. Variables' names are self-explaining. See the main text for fuller descriptions.

Table 4.A6. Correlation between net ODA and aid and monthly rainfalls

VARIABLES	OLS ΔAid_t
$Rainfall_t$	344.777*
	(189.410)
Aid_{t-1}	-0.000
	(0.000)
Observations	55
R-squared	0.040

Notes: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Estimation is:

$$\Delta Aid_t = Rainfall_t + Aid_{t-1} + \epsilon_t$$

where Aid_t is the net official development assistance and aid (constant 2013 US '000 dollar) to Uganda in year t according to the World Bank for years from 1960 – 2015. $\Delta Aid_t = Aid_t - Aid_{t-1}$, $Rainfall_t$ is the yearly average of the precipitation indicators for Uganda, obtained from the Climate Change Knowledge Portal (The World Bank Group). Despite descriptive, the result shows a clear significant correlation between the rainfall average and aid flows into Uganda. It addresses the concern that the 2007 surge in aid was in accordance with some political cycle instead of due to the shock in rainfall.

Table 4.A7. Robustness to excluding NGOs whose managers work for many NGOs

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Outside NGOs ≤ 10			Outside NGOs ≤ 15			Outside NGOs ≤ 100		
	Panel A: Using the count index as the dependent variable								
Incentives	0.11	-2.38**	-2.10**	0.11	-2.42**	-2.15**	0.10	-2.43**	-2.15**
	(0.23)	(1.06)	(1.09)	(0.23)	(1.07)	(1.11)	(0.23)	(1.07)	(1.11)
Observations	384	384	275	386	386	277	390	390	277
	Panel B: Using the HHI index as the dependent variable								
Incentives	-0.02	-0.15	-0.17	-0.02	-0.15	-0.16	-0.02	-0.18	-0.20
	(0.03)	(0.15)	(0.16)	(0.03)	(0.15)	(0.16)	(0.03)	(0.16)	(0.16)
Observations	317	317	225	319	319	227	322	322	229
Estimators	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample for affected areas	Full	Full	Restricted	Full	Full	Restricted	Full	Full	Restricted

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Estimates are multiplied by 100 for ease of interpretation. This table re-estimates our analysis when we exclude NGOs whose managers involved in more than 10 (15, 100) outside organisations. Out of 391 NGOs reported in the main analysis, there are respectively 7, 5, 1 NGOs whose managers involved in more than 10, 15, 100 outside organisations. The samples used for affected areas are either full (including all available NGOs) or restricted (including NGOs most vulnerable to an extreme flood during the period 1988 – 2017). All estimations are with the list of controls specified in the main text.

4.6.6. Using clustered errors with a small number of clusters and uneven sizes

This section reports confidence intervals of our estimates using pairs clustered bootstrap-based procedure suggested by Esarey and Menger (2018), Ibragimov and Muller (2010), and Cameron et al. (2008) to obtain accurate inference about the significance when the data is clustered at a small number of clusters with uneven sizes. We manually calculate the IV estimates, then calculate the standard errors by performing pairs bootstrapping at 500 replications by sampling all observations with replacement by cluster groups (22 groups). We report 95% confidence intervals for the main estimates in Table 4.A8. Overall, all confidence intervals of the main variable of interest (INCENTIVES) do not contain zeros, supporting our inference on the statistical significance of our preferred estimations.

Table 4.A8. Robustness to using pairs clustered bootstrapped procedure

Variables	Dependent variable: Count Index			
	Panel A. 2SLS results for full sample (N = 391)			
	Coefficient	Prob>t	95%_CI_low	95%_CI_high
INCENTIVES	-2.53	0.00	-4.26	-0.79
KAMPALA	-56.54	0.00	-92.18	-20.90
CHANGED FOCUS	35.56	0.00	4.65	66.46
DISTRICT	4.32	0.00	1.12	7.52
GEOGRAPHICAL EXPANSION	45.26	0.00	6.65	83.88
LOGSTAFF	26.04	0.00	5.40	46.68
VOTE_ACTIVITY	-47.67	10	-104.26	8.93
RELIGIOUS	-14.48	50	-57.57	28.61
TENURE	10.69	0.00	4.60	16.78
TENURE_SQUARED	-0.31	0.00	-0.50	-0.12
OTHER_NGOS	-2.23	0.00	-3.19	-1.27
CONSTANT	467.85	0.00	327.17	608.53
	Panel B. 2SLS results for restricted sample (N = 323)			
INCENTIVES	-2.47	0.00	-4.54	-0.40
KAMPALA	-50.04	0.00	-90.18	-9.89
CHANGED FOCUS	38.54	0.00	2.75	74.32
DISTRICT	4.73	0.00	2.20	7.27
GEOGRAPHICAL EXPANSION	33.19	10	-14.28	80.66
LOGSTAFF	30.09	10	-0.36	60.55
VOTE_ACTIVITY	-54.43	10	-127.12	18.25
RELIGIOUS	12.15	60	-57.72	82.01
TENURE	10.01	0.00	4.68	15.34
TENURE_SQUARED	-0.26	0.00	-0.38	-0.15
OTHER_NGOS	-0.47	80	-4.34	3.40
CONSTANT	447.02	0.00	278.72	615.31

Note. Pairs clustered bootstrapped-based standard errors (at 500 replications). Estimates are multiplied by 100 for ease of interpretation. The samples used are either full (including all available NGOs) or restricted (including NGOs vulnerable to an extreme flood during the period 1988 – 2017).

Chapter 5. Conclusion

The three papers present interconnected attempts to understand the factors underlying the behaviour and governance aspects of non-profit organisations around the world. The pieces of evidence in this thesis are robust and consistent with a range of assumptions and estimators. This chapter summarises the key findings of the papers and discusses limitations and potential steps to push forward this research agenda.

Chapter 2 provides the first systematic study on the reporting behaviour of non-profit organisations in Britain, a well-developed market for non-profits and charities. We use a digital analysis based on Benford's Law to measure the degree of misreporting in financial data submitted by a large, nationally representative sample of British charities. We find statistical evidence that financial figures from 25% of the charities collected in the UK Third Sector Data during the period 2008-2015 do not conform to Benford's Law at the 5% significance level, suggesting potential misreporting. We also connect this new measure of misreporting to two popular metrics of non-profit performance, namely programme ratio (the percentage of total income spent on charitable activities) and overhead ratio (the percentage of total income spent on governance activities). We find evidence to support the belief among non-profits that over-spending on governance activities and back-offices could be counter-productive by failing to motivate the organisations to adhere to the accountability norm. NPOs' charitable effort and the oversight mechanisms, measured by spending on charitable activities and governance activities respectively, need to be sufficiently high (70% and 15% of total income) so that their public financial data could become credible. We show that both programme ratios and overhead ratios could provide misleading signals about non-profit information transparency and quality. Although our thresholds should be interpreted as indicative, our work provides an implementable method to identify relevant indicators to assist regulators and donors when assessing financial reports and giving support to overhead activities.

Another limitation of the analysis in Chapter 2 is that due to data availability and the survey structure, we are not able to sufficiently address the concern of information withholding. That is, charities could refuse to report unfavourable information to the survey enumerators. We attempt to check and remedy the concern with a Heckman correction model and alternative methods for dealing with the number of financial items used in constructing our measures of misreporting. Nevertheless, a more in-depth analysis would be needed to shed light on how the withholding behaviour could impact our results.

Chapter 3 addresses this concern and extends the analysis to distinguish between the decision to withhold information and the decision to report accurately using data

available for Uganda. We find evidence that the underlying mechanism for nondisclosure of financial accounts and the provision of false information to be uncorrelated for a nationally representative sample of Ugandan development NGOs in the early 2000s. Instead, the shortage of skills and resources for bookkeeping tasks appear to be the main reasons for the unavailability of some core financial information. As Chapter 3 speaks to a development NGO sector in its early days, interpretations and their external validity for Chapter 2's British context could be limited. Nonetheless, together with the evidence on the need for a balance between programme ratio and overhead ratio, we advocate for the necessity to mobilise the donor community to contribute towards funding oversight mechanisms. This funding initiative could help harmonise the accountability pressure on organisations with limited-resources in both the developed and developing areas. The funding is crucial when under-staffed NGOs are unable to accommodate both their main services and expensive bookkeeping procedures. Designing specific financial packages for overhead costs could be the first step to improve the accountability of the sector and maintain the benefits that come with grassroots NGOs.

Chapter 3 also points to an alignment between higher community-based performance evaluation scores and accurate reporting behaviour by Ugandan development NGOs. Given the alignment and the increased emphasis on community (client) based assessment, collecting assessments directly from the beneficiaries could be more efficient than requiring self-reported information from NGOs. For one, doing so would avoid the organisations having to divert their scarce resources away from main charitable missions. For another, involving communities in monitoring project-implementation by organisations could potentially enhance the accountability of the aid sector, particularly grassroots NGOs. Given the importance of transparency for good governance in the NGO sector, research on this topic remains limited and little is known on how to access the benefits of community-based monitoring. There is a need for further data collection on a larger scale to address the potential concerns over the small sample size in the paper.

Chapters 2 and 3 advocate and demonstrate the feasibility of Benford's Law to detect and measure irregularities in financial accounts of non-profit organisations. As for any other measures of information irregularities and corruption, the method is neither definitive nor completely fool-proof. The premise underlying the method also remains crucial to its applicability. That is, without manipulations financial and accounting data should follow Benford's Law. We assume that deviations farther away from the theoretical Benford distribution indicate higher degrees of irregularities in the observed distribution, following a range of psychology and statistics studies. Although these assumptions are plausible and have been widely used, statistical research is needed to establish a standard set of conditions for which Benford's Law could best apply. Under these assumptions, the

method is an efficient and informative approach to flagging potential misreporting by any organisation that could subsequently be targeted for auditing. As such, our papers could be a significant contribution to reduce the monitoring and evaluation costs of the regulatory bodies for the ever-growing non-profit sector around the world.

Chapter 4 provides a piece of behavioural evidence on what motivates development NGOs in Uganda. Here, we examine an understudied but important behaviour of development NGOs, namely the decision to diversify their activities. It is important because although activity diversification could accommodate a wider range of beneficiaries, doing so could lead to inefficiency and divergence away from the core missions that the donors care for and the intended beneficiaries truly need. We predict that a higher incentive, measured as increased contractual funding, will decrease diversification for NGOs motivated by reducing uncertainty; while a higher incentive will increase diversification for NGOs motivated by private benefits. Exploiting variations in funding after the surge of international grants toward several Ugandan regions following an unexpected flood in 2007, we show that NGOs respond to an increase in the proportion of incomes from contracts (such as grants, membership or user fees) by lowering the number of activities they offer. According to our model, this empirical evidence suggests that Ugandan NGOs, on average, consider risk reduction as the primary motivation for the diversification decision rather than to self-benefit from the behaviour. The implication is that the NGO management might not consider activity diversification as desirable but a strategy for survival. Policies to ensure financial stability could help grassroots NGOs focus on their main charitable missions rather than expanding to business activities for income. This stability would improve the efficiency of the development NGO sector, particularly in developing countries, when NGOs are allowed to perform their primary tasks. Since our analysis relies on a principal-agent model to guide the empirical specification and inference, our interpretations are restricted to the set of (plausible) underlying assumptions. Our instrumental variable strategy only gives a local inference, that is, the results are uninformative to NGOs who would have (or not) received increased contractual funding regardless of the flood in 2007. future studies using other empirical strategies such as experimental data and calibration methods would help clarify the causal link between behavioural motivators and NGO strategies as well as the dynamic interactions between NGOs and their donors. This agenda is an exciting and growing strand of the literature.

Taken together, the essays provide a set of suggestive and causal evidence to help the public, donors, and regulators in assessing non-profits and designing aid packages to incentivise them, particularly in the development context where grassroots organisations often lack financial stability. This implies a benefit from open dialogues between practitioners and regulators about the daily decisions of organisations in the development

sector, from administrative tasks to operational strategies such as expanding activities. Non-profits could benefit from these discussions in shaping their policies and management toward effective governance mechanisms and transparent communications to both their funders and their beneficiaries.

References

(Dataset) Alcock, P. & Mohan, J. (2015). *Third Sector Research Centre research data collection*. Colchester, Essex: Economic and Social Research Council. https://dx.doi.org/10.5255/UKDA-SN-850933.

Aaker, J., Vohs, K. D., & Mogilner, C. (2010). Nonprofits are seen as warm and for-profits as competent: Firm stereotypes matter. *Journal of Consumer Research*, 37(2), 224-237.

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.

Aggarwal, R. K., & Samwick, A. A. (2003). Why do Managers Diversify their Firms? Agency Reconsidered. *Journal of Finance*, *58*(1), 71–118. http://doi.org/10.1111/1540-6261.00519

Aldashev, G., & Navarra, C. (2018). Development NGOs: Basic Facts. *Annals of Public and Cooperative Economics*, *89*(1), 125–155. http://doi.org/10.1111/apce.12188

Aldashev, G., & Verdier, T. (2010). Goodwill bazaar: NGO competition and giving to development. *Journal of Development Economics*, *91*(1), 48-63

Aldashev, G., Jaimovich, E., & Verdier, T. (2018). Small is beautiful: Motivational allocation in the nonprofit sector. *Journal of the European Economic Association*, *16*(3), 730–780. http://doi.org/10.1093/jeea/jvx024

Aldashev, G., Marini, M., & Verdier, T. (2015). Governance of non-profit and non-governmental organisations–within-and between-organisation analyses: an introduction. *Annals of Public and Cooperative Economics*, *86*(1), 1-5.

Almond, D., & Xia, X. (2017). Do nonprofits manipulate investment returns? *Economics Letters*, *155*, 62–66. http://doi.org/10.1016/j.econlet.2017.03.014

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, *24*(1–2), 3–61. http://doi.org/10.1016/0304-4076(84)90074-5

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.

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy*, *97*(6), 1447–1458. http://doi.org/10.1086/261662.

Arikan, A. M., & Stulz, R. M. (2016). Corporate Acquisitions, Diversification, and the Firm's Life Cycle. *The Journal of Finance*, *71*(1), 139–194.

http://doi.org/10.1111/jofi.12362.

Asatryan, Z., Bittschi, B., & Doerrenberg, P. (2017). Remittances and Public Finances: Evidence from Oil-Price Shocks. *Journal of Public Economics*.

Auriol, E., & Brilon, S. (2018). Nonprofits in the Field : An Economic Analysis of Peer Monitoring and Sabotage. *Annals of Public and Cooperative Economics*, *89*(1), 157–174. http://doi.org/10.1111/apce.12197

Avis, E., Ferraz, C., & Finan, F. (2016). *Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians*. Cambridge, MA. Retrieved from http://www.nber.org/papers/w22443.pdf

Banks, N., Hulme, D., & Edwards, M. (2015). NGOs, States, and Donors Revisited: Still Too Close for Comfort? *World Development*, *66*, 707–718. http://doi.org/10.1016/J.WORLDDEV.2014.09.028

Barabesi, L., Cerasa, A., Cerioli, A., & Perrotta, D. (2018). Goodness-of-Fit Testing for the Newcomb-Benford Law With Application to the Detection of Customs Fraud. *Journal of Business and Economic Statistics*, *36*(2), 346–358. http://doi.org/10.1080/07350015.2016.1172014

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., & Owens, T. (2004). The Governance of Non-Governmental Organizations. *World Development*, *44*(0), 657–679. Retrieved from http://www.sciencedirect.com/science/article/pii/S0305750X05000082

Barr, A., Fafchamps, M., & Owens, T. (2005). The Governance of Non-governmental Organisations in Uganda. *World Development*, 33(4), 657-679.

Baum, C. F., Lewbel, A., Schaffer, M. E., & Talavera, O. (2012). Instrumental variables estimation using heteroskedasticity-based instruments. In *United Kingdom Stata User's Group Meetings* (Vol. 7).

Baum, C. F., Schaffer, M. E., & Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal*, *3*(1), 1-31.

Benabou, R., & Laroque, G. (1992). Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility. *The Quarterly Journal of Economics*, *107*(3), 921–958. http://doi.org/10.2307/2118369

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*(5), 1652–1678. http://doi.org/10.1257/aer.96.5.1652

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical society*, 551-572.

Besley, T., & Ghatak, M. (2005). Competition and Incentives with Motivated Agents. *American Economic Review*, *95*(3), 616–636.

http://doi.org/10.1257/0002828054201413

Bettin, G., Lucchetti, R., & Zazzaro, A. (2012). Endogeneity and sample selection in a model for remittances. *Journal of Development Economics*, *99*(2), 370–384. http://doi.org/10.1016/j.jdeveco.2012.05.001

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.

Beyer, A., Guttman, I., & Marinovic, I. (2014). Optimal contracts with performance manipulation. *Journal of Accounting Research*, *52*(4), 817–847. http://doi.org/10.1111/1475-679X.12058

Bhattacharya, R., & Tinkelman, D. (2009). How tough are better business bureau/wise giving alliance financial standards?. *Nonprofit and Voluntary Sector Quarterly*, *38*(3), 467-489.

Björkman-Nyqvist, M. (2013). Income shocks and gender gaps in education: Evidence from Uganda. *Journal of Development Economics*, *105*, 237-253.

Boland, P. J., & Hutchinson, K. (2000). Student selection of random digits. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*(4), 519-529.

Boris, E. T., & Steuerle, C. E. (2006). *Nonprofits and Government: Collaboration and Conflict*. Urban Institute Press.

Boyle, J. (1994). An application of Fourier series to the most significant digit problem. *The American Mathematical Monthly*, *101*(9), 879-886.

Brass, J. N. (2012). Why do NGOs go where they go? Evidence from Kenya. *World Development*, *40*(2), 387-401.

Burger, R., & Owens, T. (2010). Promoting transparency in the NGO sector: Examining the availability and reliability of self-reported data. *World Development*, *38*(9), 1263–1277. http://doi.org/10.1016/j.worlddev.2009.12.018

Burger, R., & Owens, T. (2013). Receive grants or perish? The survival prospects of Ugandan Non-Governmental Organisations. *The Journal of Development Studies*, 49(9), 1284-1298.

Burgstahler, D., & Dichev, 2. (1997). Earnings Management to Avoid Earnings Decreases and Losses. *Journal of Accounting and Economics*, *24*(1), 99-126.

Burns, N., & Kedia, S. (2006). The impact of performance-based compensation on misreporting. *Journal of Financial Economics*, *79*(1), 35–67. http://doi.org/10.1016/j.jfineco.2004.12.003

Caliendo, M., Lee, W. S., & Mahlstedt, R. (2017). The gender wage gap and the role of reservation wages: New evidence for unemployed workers. *Journal of Economic Behavior & Organization*, *136*, 161-173.

Camerer, C. F. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction*, New York, NY: Russell Sage Foundation and Princeton University Press.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317-372.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414-427.

Campa, J. M., & Kedia, S. (2002). Explaining the diversification discount. *The journal of finance*, *57*(4), 1731-1762.

Carpenter, J., & Gong, E. (2016). Motivating Agents: How much does the mission matter?. *Journal of Labor Economics*, *34*(1), 211-236.

Carroll, D. A., & Stater, K. J. (2009). Revenue diversification in nonprofit organizations: Does it lead to financial stability. *Journal of Public Administration Research and Theory*, 19(4), 947-966.

Carroll, G. (2015). Robustness and linear contracts. *American Economic Review*, *105*(2), 536-63.

Chaney, P. K., Faccio, M., & Parsley, D. (2011). The quality of accounting information in politically connected firms. *Journal of Accounting and Economics*, *51*(1–2), 58–76. http://doi.org/10.1016/j.jacceco.2010.07.003

Chen, Q. (2016). Director Monitoring of Expense Misreporting in Nonprofit Organizations: The Effects of Expense Disclosure Transparency, Donor Evaluation Focus and Organization Performance. *Contemporary Accounting Research*, *33*(4), 1601–1624. http://doi.org/10.1111/1911-3846.12218

Cho, W. K. T., & Gaines, B. J. (2007). Breaking the (Benford) law: Statistical fraud detection in campaign finance. *American Statistician*, *61*(3), 218–223. http://doi.org/10.1198/000313007X223496

Cormier, D., Magnan, M., & Van Velthoven, B. (2005). Environmental disclosure quality in large German companies: Economic incentives, public pressures or institutional conditions? *European Accounting Review*, *14*(1), 3–39. http://doi.org/10.1080/0963818042000339617

Cragg, J. G. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica*, *39*(5), 829. http://doi.org/10.2307/1909582

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 1431-1451.

Crocker, K. J., & Slemrod, J. (2007). The economics of earnings manipulation and managerial compensation. *The RAND Journal of Economics*, *38*(3), 698-713.

Davidson, R., & MacKinnon, J. G. (1981). Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica*, *49*(3), 781–793. http://doi.org/10.2307/1911522

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*(2–3), 344–401. http://doi.org/10.1016/j.jacceco.2010.09.001

Deininger, K., & Mpuga, P. (2005). Does greater accountability improve the quality of public service delivery? Evidence from Uganda. *World Development*, *33*(1), 171–191. http://doi.org/10.1016/j.worlddev.2004.09.002

DellaVigna, S., & Pope, D. (2017). What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies*. (Corrected Proof) https://doi.org/10.1093/restud/rdx033.

Dixit, A. (2002). Incentives and Organizations in the Public Sector: An Interpretative Review. *The Journal of Human Resources*, *37*(4), 696. http://doi.org/10.2307/3069614

Dow, W. H., & Norton, E. C. (2003). Choosing between and interpreting the Heckit and two-part models for corner solutions. *Health Services and Outcomes Research Methodology*, *4*(1), 5–18. http://doi.org/10.1023/A:1025827426320

Dreher, A., Nunnenkamp, P., Öhler, H., Weisser, J. (2012). Financial Dependence and Aid Allocation by Swiss NGOs: A Panel Tobit Analysis. *Economic Development and Cultural Change*, 60(4), 829-867.

Durtschi, C., & Easton, P. (2005). Earnings management? The shapes of the frequency distributions of earnings metrics are not evidence ipso facto. *Journal of Accounting Research*, *43*(4), 557-592.

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*, *99*(99), 17–34. http://doi.org/DOI:

Dye, R. A. (1986). Proprietary and Nonproprietary Disclosures. *The Journal of Business*, *59*(2), 331. http://doi.org/10.1086/296331

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. http://doi.org/10.1016/S0305-750X(03)00014-7

Edwards, M., & Hulme, D. (1992). Scaling up NGO impact on development: learning from experience. *Development in Practice*, 2(2), 77-91.

Emran, M. S., & Hou, Z. (2013). Access to markets and rural poverty: evidence from household consumption in China. *Review of Economics and Statistics*, *95*(2), 682-697.

Epstein, M. J., & Yuthas, K. (2014). Measuring and Improving Social Impacts. In *Measuring and Improving Social Impacts: A Guide for Nonprofits, Companies and Impact Investors* (Vol. 44, pp. 0–9). Routledge. http://doi.org/10.4324/9781351276245.

Erel, I., Jang, Y., & Weisbach, M. S. (2015). Do acquisitions relieve target firms' financial constraints?. *The Journal of Finance*, *70*(1), 289-328.

Fang, H., & Gong, Q. (2017). Detecting potential overbilling in medicare reimbursement via hours worked. *American Economic Review* (Vol. 107, pp. 562–591). http://doi.org/10.1257/aer.20160349

FAO. (2008). *Special report: Assessment of the impact of 2007 flood on food and agriculture in Eastern and Northern Uganda.* Accessed at

Fehr, E., & Schmidt, K. M. (2006). Chapter 8 The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories (pp. 615–691). http://doi.org/10.1016/S1574-0714(06)01008-6

Fruttero, A., & Gauri, V. (2005). The Strategic Choices of NGOs: Location Decisions in Rural Bangladesh1. *Journal of Development Studies*, *41*(5), 759–787. http://doi.org/10.1080/00220380500145289

Fudenberg, D., & Tirole, J. (1995). A Theory of Income and Dividend Smoothing Based on Incumbency Rents. *Journal of Political Economy*, *103*(1), 75–93. http://doi.org/10.1086/261976

Gabaix, X. (1999). Zipf's law for cities: an explanation. *Quarterly Journal of Economics*, 114(3), 739-767.

Garven, S. A., Hofmann, M. A., & McSwain, D. N. (2016). Playing the Numbers Game. *Nonprofit Management and Leadership*, *26*(4), 401-416.

Gibbons, R., & Murphy, K. J. (1992). Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence. *Journal of Political Economy*, *100*(3), 468–505. http://doi.org/10.1086/261826

Gilliam, T. A., Heflin, F., & Paterson, J. S. (2015). Evidence that the zero-earnings discontinuity has disappeared. *Journal of Accounting and Economics*, *60*(1), 117-132.

Gneezy, U. (2005). Deception: The Role of Consequences. *American Economic Review*, *95*(1), 384–394. http://doi.org/10.1257/0002828053828662

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, *25*(4), 191-209.

Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal*

of Economic Behavior and Organization, *93*, 293–300. http://doi.org/10.1016/j.jebo.2013.03.025

Goldman, E., & Slezak, S. L. (2006). An equilibrium model of incentive contracts in the presence of information manipulation. *Journal of Financial Economics*, *80*(3), 603-626.

Grendar, M., Judge, G., & Schechter, L. (2007). An empirical non-parametric likelihood family of data-based Benford-like distributions. *Physica A: Statistical Mechanics and its Applications*, *380*, 429-438.

Hansen, C., Hausman, J., & Newey, W. (2008). Estimation with Many Instrumental Variables. *Journal of Business & Economic Statistics*, *26*(4), 398-422.

Hatte, S., & Koenig, P. (n.d.). The Geography of NGO Activism against Multinational Corporations. *The World Bank Economic Review*. http://doi.org/10.1093/wber/lhy007

Hausman, J. (2001). Mismeasured variables in econometric analysis: problems from the right and problems from the left. *The Journal of Economic Perspectives*, *15*(4), 57-67.

Hayashi, F. (2000). Econometrics. Princeton. *New Jersey, USA: Princeton University*.

Healy, P. M. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics*, *7*(1–3), 85–107. http://doi.org/10.1016/0165-4101(85)90029-1

Herzlinger, R. E. (1996). Can public trust in nonprofits and governments be restored? *Harvard Business Review*, *74*(2), 97–107. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10156469

Hill, T. P. (1998). The First Digit Phenomenon. *American Scientist*. http://doi.org/10.1511/1998.31.815

Hodge, M. M., & Piccolo, R. F. (2005). Funding source, board involvement techniques, and financial vulnerability in nonprofit organizations: A test of resource dependence. *Nonprofit Management and Leadership*, *16*(2), 171–190. http://doi.org/10.1002/nml.99

Hodge, M. M., & Piccolo, R. F. (2005). Funding source, board involvement techniques, and financial vulnerability in nonprofit organizations: A test of resource dependence. *Nonprofit Management and Leadership*, *16*(2), 171–190. http://doi.org/10.1002/nml.99.

Hofmann, M. A., & McSwain, D. (2013). Financial disclosure management in the nonprofit sector: A framework for past and future research. *Journal of Accounting Literature*, *32*(1), 61–87. http://doi.org/10.1016/j.acclit.2013.10.003

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: on the affective psychology of value. *Journal of Experimental Psychology: General*, *133*(1), 23.

Hurkens, S., & Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics*, *12*(2), 180–192. http://doi.org/10.1007/s10683-008-9208-2

Ibragimov, R., & Müller, U. K. (2010). t-Statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics*, *28*(4), 453-468.

Imas, A. (2014). Working for the "warm glow": On the benefits and limits of prosocial incentives. *Journal of Public Economics*, *114*, 14–18. http://doi.org/10.1016/j.jpubeco.2013.11.006.

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475

Jacob, B. A., & Levitt, S. D. (2003). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *The Quarterly Journal of Economics*, 843-877

Jensen, M. C., & Meckling, W. H. (1979). Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure (pp. 163–231). Springer, Dordrecht. http://doi.org/10.1007/978-94-009-9257-3_8

Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of accounting research*, 193-228.

Judge, G., & Schechter, L. (2009). Detecting Problems in Survey Data using Benford's Law. *Journal of Human Resources*, *44*(1), 1–24. http://doi.org/10.1353/jhr.2009.0010

Keating, E. K., & Frumkin, P. (2003). Nonprofit Reengineering Toward Accountability : a Financial More Reliable Foundation for Regulation. *Public Administration Review*, *63*(1), 3–15. http://doi.org/10.2307/977517

Keating, E. K., Fischer, M., Gordon, T. P., & Greenlee, J. S. (2005). *Assessing financial vulnerability in the nonprofit sector*. Faculty Research Working Paper Series: Hauser Center for Nonprofit Organizations, Paper no. 27, Harvard University.

Kistruck, G. M., Qureshi, I., & Beamish, P. W. (2013). Geographic and product diversification in charitable organizations. *Journal of Management*, *39*(2), 496-530.

Kossovsky, A. E. (2015). *Benford's Law*. Singapore: World Scientific Publishing.

Kothari, S. P., Shu, S., & Wysocki, P. D. (2009). Do managers withhold bad news. *Journal of Accounting Research*, *47*(1), 241–276. http://doi.org/10.1111/j.1475-679X.2008.00318.x

Krishnan, R., & Yetman, M. H. (2011). Institutional drivers of reporting decisions in nonprofit hospitals. *Journal of Accounting Research*, *49*(4), 1001–1039. http://doi.org/10.1111/j.1475-679X.2011.00413.x

Lang, M., & Lundholm, R. (1993). Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of accounting research*, 246-271.

Lee, D. S., & Lemieuxa, T. (2010). Regression discontinuity designs in economics. *Journal of Economic L iterature*, *48*(2), 281-355.

Levine, D. K. (1998). Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*, *1*(3), 593–622. http://doi.org/10.1006/redy.1998.0023

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, *30*(1), 67-80.

Linardi, S., & McConnell, M. A. (2011). No excuses for good behavior: Volunteering and the social environment. *Journal of Public Economics*, 95(5), 445-454.

Loy, J. P., Weiss, C. R., & Glauben, T. (2016). Asymmetric cost pass-through? Empirical evidence on the role of market power, search and menu costs. *Journal of Economic Behavior & Organisation*, *123*, 184-192.

Madden, D. (2008). Sample selection versus two-part models revisited: The case of female smoking and drinking. *Journal of Health Economics*, *27*(2), 300–307. http://doi.org/10.1016/j.jhealeco.2007.07.001

Mansuri, G., & Rao, V. (2012). *Localizing Development*. The World Bank. http://doi.org/10.1596/978-0-8213-8256-1

Martens, B. (2005). Why do aid agencies exist?. *Development Policy Review*, 23(6), 643-663.

McCormick, B. (1990). (Search Readings) - A Theory of Signalling During Job Search, Employment Efficiency, and "Stigmatised" Jobs. *The Review of Economic Studies*, *57*(2), 15. http://doi.org/10.2307/2297383

McGuire, S. T., Omer, T. C., & Sharp, N. Y. (2012). The impact of religion on financial reporting irregularities. *The Accounting Review*, *87*(2), 645–673. http://doi.org/10.2308/accr-10206

Mendoza-Abarca, K. I., & Gras, D. (2017). The Performance Effects of Pursuing a Diversification Strategy by Newly Founded Nonprofit Organizations. *Journal of Management*, doi:10.1177/0149206316685854.

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. http://doi.org/10.1162/REST_a_00274

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 380-391.

Miller, S. J. (2015). *Benford's law : theory and applications*. Princeton University Press.

Millimet, D. L., & Roy, J. (2016). Empirical tests of the pollution haven hypothesis when environmental regulation is endogenous. *Journal of Applied Econometrics*, *31*(4), 652-677.

Montgomery, C. A. (1994). Corporate Diversification. *Journal of Economic Perspectives*, *8*(3), 163–178. http://doi.org/10.1257/jep.8.3.163.

Morfit, N. S. (2011). "AIDS is money": how donor preferences reconfigure local realities. *World Development*, 39(1), 64-76.

Morrow, J. (2014). Benford's law, families of distributions and a test basis. *CEP Discussion Paper*, *No 1291*, 1–26. Retrieved from http://cep.lse.ac.uk/pubs/download/dp1291.pdf

Moscarini, G. (2007). Competence implies credibility. *The American Economic Review*, 97(1), 37-63.

Murphy, K. J. (1999). Executive compensation. *Handbook of Labor Economics*, 3, 2485-2563.

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

Norton, E. C., Dow, W. H., & Do, Y. K. (2008). Specification tests for the sample selection and two-part models. *Health Services and Outcomes Research Methodology*, *8*(4), 201–208. http://doi.org/10.1007/s10742-008-0037-8

Norton, W. (2014). *Transparency Begins at Home. Why charities must state who funds them.* Centre for Policy Studies. ISBN 978-1-910627-02-0.

Nye, J., & Moul, C. (2007). The political economy of numbers: on the application of Benford's law to macroeconomic statistics. *The BE Journal of Macroeconomics*, 7(1), 1-14.

Olenick, A. J. (1988). Straight talk about a nagging problem: how to improve nonprofit management. *Nonprofit World*, 6, 31-33.

Olken, B. A. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy*, *115*(2), 200–249.

Olken, B. A., & Pande, R. (2012). Corruption in Developing Countries. *Annual Review of Economics*, *4*(1), 479–509.

Olson, M. (2009). *The Logic of Collective Action* (Vol. 124). Harvard University Press.

Parente, P. M., & Santos Silva, J. S. (2012). A cautionary note on tests of overidentifying restrictions. *Economics Letters*, *115*(2), 314-317.

Planck, M. (1901). On the law of the energy distribution in the normal spectrum. *Ann. Phys*, *4*(553), 90.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model.

Journal of Econometrics, *25*(3), 303–325.

Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics, 85*(4), 777–792.

Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, *56*(4), 931.

Salamon, L. M. (2010). Putting the civil society sector on the economic map of the world. *Annals of Public and Cooperative Economics*, *81*(2), 167-210.

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.

Santos Silva, J. M. C. S., Tenreyro, S., & Windmeijer, F. (2015). Testing Competing Models for Non-negative Data with Many Zeros. *Journal of Econometric Methods*, *4*(1), 29–46. http://doi.org/10.1515/jem-2013-0005

Schennach, S. M. (2013). Measurement error in nonlinear models – A Review. In D. Acemoglu, M. Arellano & E. Dekel (Eds.), *Advances in Economics and Econometrics.*, *3*, 296-337.

Schulter, G., Mittenecker, E., & Papousek, I. (2010). A computer program for testing and analyzing random generation behavior in normal and clinical samples: The Mittenecker pointing test. *Behavior Research Methods*, *42*(1), 333–341.

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

Scott, K. (2003). *Funding matters: The impact of Canada's new funding regime on nonprofit and voluntary organisations.* Ottawa, Ontario: Canadian Council on Social Development.

Shleifer, A., & Vishny, R. W. (1989). Management entrenchment: The case of manager-specific investments. *Journal of Financial Economics*, 25(1), 123-139.

Singh, A. (2015). Transparency is great, but not at the cost of a charity's services. Retrieved September 05, 2017, from https://www.theguardian.com/voluntary-sector-network/2015/jan/08/transparency-great-cost-charity-services.

Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, 102(2), 143-153.

Smillie, I., Helmich, H., Randel, J., & German, T. (Eds.). (2013). *Stakeholders: Government-NGO partnerships for international development*. Routledge.

Stark, O., & Falk, I. (1998). Transfers, empathy formation, and reverse transfers. *The American Economic Review*, *88*(2), 271-276.

Stein, J. C. (1989). Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior. *The Quarterly Journal of Economics*, *104*(4), 655.

Steinberg, R. (2010). Principal–agent theory and nonprofit accountability. In K. Hopt & T. Von Hippel (Eds.), *Comparative Corporate Governance of Non-Profit Organisations* (International Corporate Law and Financial Market Regulation, pp. 73-126). Cambridge: Cambridge University Press.

Steinberg, R. (2010). Principal–agent theory and nonprofit accountability. In K. Hopt & T. Von Hippel (Eds.), *Comparative Corporate Governance of Non-Profit Organizations* (International Corporate Law and Financial Market Regulation, pp. 73-126). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511712128.006

Stulz, R. (1990). Managerial discretion and optimal financing policies. *Journal of Financial Economics*, 26(1), 3-27.

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*.

Tonin, M., & Vlassopoulos, M. (2014). Corporate philanthropy and productivity: Evidence from an online real effort experiment. *Management Science*, *61*(8), 1795-1811.

Trussel, J. (2003). Assessing potential accounting manipulation: The financial characteristics of charitable organisations with higher than expected program-spending ratios. *Nonprofit and Voluntary Sector Quarterly*, *32*(4), 616-634.

Trussel, J. M., & Greenlee, J. S. (2004). A financial rating system for charitable nonprofit organisations. *Research in Governmental and Nonprofit Accounting*, *11*, 93–96.

Tuckman, H. P. (1998). Competition, commercialization, and the evolution of nonprofit organiza- tional structures. *Journal of Policy Analysis and Management*, *17*, 175-194.

UOCHA. (2006). *Humanitarian Update January 2008.* Accessed at https://reliefweb.int/report/uganda/uganda-ocha-humanitarian-update-jan-2008.[10 Oct 2017].

Van Caneghem, T. (2015). NPO Financial Statement Quality: An Empirical Analysis Based on Benford's Law. *Voluntas*, 1-24.

Vansant, B. (2016). Institutional pressures to provide social benefits and the earnings management behavior of nonprofits: Evidence from the US hospital industry. *Contemporary Accounting Research*, *33*(4), 1576-1600.

Varian, H. (1972). Benford's Law (Letters to the Editor). *The American Statistician,* 26(3): 65.

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

Villas-Boas, S. B., Fu, Q., & Judge, G. (2017). Benford's law and the FSD distribution of economic behavioral micro data. *Physica A: Statistical Mechanics and its Applications*.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, *57*(2), 307. http://doi.org/10.2307/1912557

Watkins, S. C., Swidler, A., & Hannan, T. (2012). Outsourcing social transformation: Development NGOs as organisations. *Annual Review of Sociology*, *38*.

Wedig, G. J. (1994). Risk, leverage, donations and dividends-in-kind: A theory of nonprofit financial behavior. *International Review of Economics and Finance*. http://doi.org/10.1016/1059-0560(94)90011-6

Weisbrod, B. A. (1998). The nonprofit mission and its financing. *Journal of Policy Analysis and Management*, *17*, 165-174.

Woodwell, H.W., & Bartczak, L. (2008). *Is Grantmaking Getting Smart? A National Study of Philanthropic Practice.* Washington, D.C.: Grantmakers for Effective Organisations.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. *MIT Press: Cambridge, MA, USA*, 1064. http://doi.org/10.1515/humr.2003.021

Yetman, M. H., & Yetman, R. J. (2012). The effects of governance on the accuracy of charitable expenses reported by nonprofit organisations. *Contemporary Accounting Research*, *29*(3), 738-767.

Young, D. R. (2009). How nonprofit organizations manage risk. In: Destefanis, S. and Musella, M., eds. *Paid and Unpaid Labour in the Social Economy*. AIEL Series in Labour Economics, Physica-Verlag Heidelberg, pp. 33-46.

Zitzewitz, E. (2012). Forensic Economics. *Journal of Economic Literature*, *50*(3), 731–769.