Does transparency come at the cost of charitable services?

Evidence from investigating British charities[1]

Canh Thien Dang

Department of Economics,

London School of Economics, UK

c.t.dang@lse.ac.uk

Trudy Owens

School of Economics,

University of Nottingham, UK

Trudy.Owens@nottingham.ac.uk

Abstract

Recent high-profile scandals related to misuse of funding and donations have raised the demand for scrutiny over financial transparency and operational activities of non-profit organizations in developed countries. Our analysis challenges the common practice in the sector of using programme ratios and overhead costs as indicators for non-profit accountability. Using Benford's Law to measure irregularities in financial data for a large sample of public charities we estimate that 25% of the sample potentially misreport their financial information. We show theoretically and empirically that charities with a higher programme ratio (their level of spending on charitable activities), will be less likely to misreport their financial information only when their overhead costs (spending on governing activities) are also sufficiently high. Tighter monitoring becomes ineffective in increasing the sectoral transparency and accountability unless accompanied by a sufficiently high level of charitable spending.

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

(Asheem Singh, Director of the Association of Chief Executives of Voluntary Organisations,

The Guardian, 2015)

1   Introduction

Misrepresenting financial information has become a serious concern in the non-profit sector in developed countries.[1] The disconnection between the funders and the end beneficiaries has given non-profits ample opportunities and incentives to manipulate their financial reports to mislead potential donors, rating agencies, and the public about their efficiency in accomplishing philanthropic missions. 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 (such as higher spending on charitable activities) report their financial data more accurately, and whether increased governance-related overhead costs (another commonly used performance metric) correlates with a lower level of financial misinformation.

We first explore an alternative method of measuring performance based on Benford's Law to quantify misreporting in financial data for public charities in the UK.[2] The intuition is that the observed distribution of the first digits of these figures is expected to follow a theoretical distribution known as the Benford distribution.[3,4] The further the observed distribution deviates from the Benford distribution, the more likely the financial figures in the financial statement include non-random errors, either due to manipulation or human error. Using a large public dataset of over 10,000 NPOs in the UK, our preferred test finds that nearly 25% of the sample potentially misreported their financial accounts during the period 2007-2015. We then cross-check our measures with reports by the Charity Commission, the governing body of charities in England and Wales. Our analysis correctly identifies 90% of the charities recently (2016) charged for misusing their funding and tinkering with data.

The approach is an appealing alternative to the current measures of misreporting, such as

---

[1] Chen (2016) documents scandals of fund misappropriation, abuse of power and lack of transparency in the non-profit sectors. Norton (2014) suggests that the financial figures of the 50 largest UK charities by income could be more than double the self-reported accounts. Keating et al. (2008) find 74% of the regulatory filings from American non-profits fail to properly report categorical expenses.

[2] There has been a revised interest in Benford's Law, see Amiram et al (2015) and Villas-Boas et al. (2017).

[3] One order of magnitude means one value is about ten times different in quantity than the other.

[4] 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).

spending on charitable activities and/or overhead costs, for several reasons. First, it is highly replicable, and simple to calculate. Second, the required financial data are already in the public domain. Third, unlike previous studies using the distributional properties of financial figures in the non-profit literature, we can construct a misinformation measure for each charity, allowing a between-organisation analysis. Finally, it is acknowledged that people are rarely capable of manipulating datasets that follow either Benford's Law or a random distribution (see Schulter et al., 2010; Boland and Hutchinson, 2000). Like other measures in "forensic economics" (Zitzewitz, 2012), the approach is not fail-proof, nor will it substitute for full auditing. Nevertheless, we believe it can serve as a useful and relatively low-cost first step for effective and more targeted audits.

We then examine the link between the two popular performance metrics used in the charity sector and the organisational accountability literature which remain understudied by academics and poorly understood by policymakers (Aldashev et. al, 2015). Hampered by a lack of systematic organisational data and the complexity of annual reports (Dechow et al., 2010), policymakers and donors interested in accountability have to date had to rely on statistics such as programme ratios (the proportion of the total income spent on charitable activities) and overhead costs (the proportion of the total income spent on administrative activities). Garven et al.'s (2016) review of recent surveys of American private donors and charity watchdogs finds that programme ratios are a significant factor in both the donation decisions of donors, and regulatory policies of charity rating agencies and other non-profit governing boards. The consensus among individuals and donors surveyed is that approximately 75% of the total budget should be spent on charitable activities; while the American Institute of Philanthropy suggests a "reasonable" ratio of 60% or higher. Aebischer (2012) reports the maximum overhead costs supported by European institutional donors ranges from 10 to 20%; while Charity Navigator, an American charity evaluator, purports 25% of the budget spent to be spent on overheads as a strong performance indicator (cited in Exley, 2019). Using experiments, Exley (2019) and Gneezy et al. (2014) respectively show that high programme ratios and low overhead costs facilitate donations and are used as an excuse for donation refusal. Yet, despite the popularity of these metrics there has been little evidence or assessments of their usefulness (Exley, 2019; Garven et al., 2016).

In this paper, we cast doubt on the belief that increased programme ratios and governance-related overhead costs alone would necessarily improve accountability. Keeping accurate records is costly and requires financial and human resources that could otherwise be spent on charitable activities. The opening quote by Singh (2015) asserts this strongly held

view from within the sector: while transparency is desired, spending on charitable services needs to remain the priority. Nor is there conclusive evidence of the effectiveness of increased accountability, even in the corporate literature (Hermalin & Weisbach, 2012). Without demonstrated value, increased spending on governance could incur unnecessary costs and distort charitable agendas (Aldashev et al., 2015). We build a simple a theoretical model which shows that, alone, neither higher charitable effort, nor stricter governance necessarily induces less bias in financial reporting. Instead, the agent (charity) could strategically misreport due to the trade-off between charitable effort and accuracy in reporting.

We test the predictions by examining how the level of misreporting, measured by our Benford digital analysis, is affected by (i) the proportion of income spent on charitable activities (or programme ratio), a proxy for the charitable effort, and (ii) the proportion of income spent on governance activities (or governance-related overhead costs), a proxy for the oversight mechanism. To address the concerns of potential mismeasurement errors and confounders that affect both the tendency to exert effort and misreport, we extend the linear regressions by employing two instrumental variables strategies. The first uses two instruments that are easily observed or easy to cross-check by interested parties, namely, the number of official staff and spending on social security. The second strategy utilizes a novel estimator by Lewbel (2012) that uses conditional second moments to construct in-data instruments that do not require the standard exclusion restriction. Results from linear regressions controlling for various characteristics and the two IV estimations, with a battery of robustness checks, provide a consistent pattern supporting an interaction effect. We find, first, NPOs with higher charitable spending report more accurately only when governance spending exceeds a certain level (15% of total income in our preferred specification). To put the number in perspective, only 7% of the charities in our sample spend this figure of 15% highlighting the lack of resources available for these tasks. Woodwell and Bartczak's (2008) survey for the US sector documents that 80% of US donors did not include sufficient overhead expenses related to reporting requirements. Second, NPOs with larger 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 maintaing accurate financial numbers are higher (for example, in larger and older charities), and lower when the probability of being detected as a data manipulator is high (such as the financial reports being audited or receiving government grants) or when the charities report more non-zero values. Taken together, we provide evidence against the naive use of programme ratios and overhead costs as indicators for non-profit transparency and accountability. Instead, we call for donors to provide a more

balanced expectation of the use of income when funding charitable activities and governance tasks if the aim is to improve the sector's effectiveness and accountability.

Our paper contributes to and bridges several fields of literature. First, it relates to "forensic economic" studies (Zitzewitz, 2012) on the prevalence and determinants of misreporting. Our application of Benford's Law is closely tied 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). We advocate the use of Benford's Law as an effective, investigative tool in the non-profit literature to flag potential charities for early investigation, particularly in a context where data is scarce. Second, we complement the for-profit literature on the determinants of information manipulation (see Bayer et al., 2010 for a review). Most of the literature suggests that firms may 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. Third, our theoretical framework provides an alternative explanation of firms and charities misbehaviour. 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). Deviating slightly, our focus is on the charitable spending and the oversight mechanism as the key explanations for charities' misreporting behaviour. Our theoretical and empirical results complement previous theoretical predictions that there exists a point beyond which additional corporate governance decreases firm value (Hermalin and Weisbach, 2012). Finally, we contribute to the growing economics and accounting literature on the usefulness and limitations of performance metrics in evaluating charity effectiveness and transparency (Exley, 2019 Garven et al., 2016). We suggest that higher programme ratios and governance-related overhead costs might fail to improve accountability and accuracy of financial reports.

The paper proceeds as follows. Section 2 explains Benford's Law, the UK Third Sector data and how our proxies are constructed. We discuss several caveats and how to remedy them in Section 2.3. We present the theoretical model in Section 3. Section 4 describes our empirical analysis. Section 5 reports the main findings. Section 6 summarises various robustness checks. Section 7 concludes.

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

2.1   Forensic Economics Studies and Benford's Law

There are several approaches to measuring misreporting in financial data. One popular method is through estimating accounting models and comparing the reported and predicted activities (Dechow et al., 2010). Another approach uses distributional properties of financial numbers such as discontinuity at zero of earnings or rates of returns to identify potential errors (Burgstahler and Dichev, 1997; Lee and Lemieux, 2010).[5] These measures have inherent drawbacks, such as, correlations with the underlying characteristics of the organisation and/or manager (see Amiram et al., 2015), or the confounding effects of scaling, sample selection and research designs (see Durtschi et al., 2005; Gilliam et al., 2015 for the discontinuity approach). They are also only feasible when detailed records or forward-looking information such as expected incomes or cashflows are available. Due to issues of data availability, non-profit studies are often constrained to evaluating categories of financial records, such as, fundraising expenses or returns on external investments, instead of more useful disaggregated data. A notable example includes Almond and Xia's (2017) study which shows that some US non-profits manipulate investment returns around zero to avoid revealing negative outcomes. Other studies look at 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).

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 discovered 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

---

[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.

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 of being the leading number). Second, if the probability distribution function of the logarithm of N is smooth and symmetric, a number will be in the interval 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% chance. 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 this 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 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 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 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 accounting papers.[7] Nigrini (2012) documents successful applications 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.

---

[6] See Villas-Boas et al. (2017) recent statistical evidence of the law's applicability to economic behavioural data.

[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.

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.8 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).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 Table A1). The variety of financial items, drawn from different types of underlying distributions (such as revenues, donations, or expenses), increases the chance of conformity to Benford's Law.[10] All the figures are recorded rather than constructed from raw data. This feature preserves the "naturalness" of the underlying distributions as it avoids both seasonal adjustments, and the use of statistical methods used in constructing an index.

For each charity, we use all the financial entries that appear on each annual report to calculate our measure of misreporting. In a standard financial statement, there are 135 entries (see Table A1). Ideally, we would use each annual report to construct organisation-year measures to study the dynamics of misreporting, however, many items are recorded as zero. Since there is no coding for missing observations, we treat these zeros as genuine information, that is, transactions whose values are zero. The zero items present a challenge to our analysis as the construction of the Benford measure requires at least 100 non-zero items in each tested unit to ensure the power of the related statistical tests.[11] To achieve the statistical power, we

---

8 Due to survey design, only non-profits with the total income of at least £250,00 are collected.

9 We first convert all 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).

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 $P_e(d_i) \geq 5$ for all $d_i \in \{1, \ldots, 9\}$. That is, the lower bound for N to be a valid statistic is $N \geq 5/P_e(d_i = 9) > 100$. Milchaski and Stoltz (2013) run

aggregate each NPO's annual reports over the available years to calculate a measure of aggregated misreporting.

The interpretation of our results is not affected by this practice. First, British charities provide their annual financial statements in blocks (in other words, several consecutive financial statements are collected at one point in time). Charities therefore can inject errors in multiple annual accounts in one submission so that it is appropriate to consider these multiple annual accounts as an aggregated dataset. Second, the practice of aggregating yearly data is not uncommon (see Amiram et al., 2015 for aggregated simulated accounting data; and Michalski and Stoltz (2013) for aggregated macroeconomic data).[12] Kossovsky (2015, p.90) provides examples of aggregated data over months/years and an industry (combining multiple companies) that also conform with the law. Finally, the aggregated financial data satisfies the two statistical conditions for an accounting dataset to conform to Benford's Law laid out in Durtschi et al. (2004), namely: (1) positive skewness and (2) mean-to-median ratio larger than one. Figure A1 in the Appendix demonstrates that our aggregated data for each charity satisfy the two conditions.

In Section 6, we split the sample into two periods (pooling four years of data together) to show that our results are not sensitive to the aggregation even when we allow for the time dynamics of misreporting. Because splitting the data severely reduces our sample, for our main analysis, we consider a cross-sectional sample of 10,322 charities that provide at least 100 non-zero financial figures over all their financial reports.

Conceptually, removing NPOs with fewer than 100 non-zero financial items after pooling could lead to selection bias. For instance, 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

---

simulations and suggest that tests for Benford's Law is powerful only for samples with at least 110-digit points. We experiment with both the cut-offs and find similar results.

[12] Amiram 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 organisation-year are expected to conform to Benford's Law. Nigrini (2011, chapter 17) uses multi-year financial statements to demonstrate the applicability of aggregate data in assessing errors and frauds by digital analysis. 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.

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 this 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 report in Appendix 8.6.4 similar results when we use a Heckman correction model to account for the possibility that some NPOs report fewer non-zero transactions to withhold information.

2.3. Measures of information misreporting

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 with 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 1):

---

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

[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).

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

where $d_i = 1, 2, ..., 9$ 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 1.[15] To interpret, the larger the *MAD statistic*, the further the deviation from the theoretical distribution under the null hypothesis that the pooled report is free of errors and misrepresentation.

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

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

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

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

---

[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.

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 $\alpha = 0.05$, $c(0.05) = 1.48$. For the *Deviate* variable, we calculate the exact p-values for K-S tests P-values by sampling from the null distribution (Monte Carlo simulation) at 10,000 replications (see Senchaudhuri et al., 1995 and Barasebi et al., 2017).

Caveats of our measures: These measures hinge on the premise that an accurate financial statement adheres to Benford's Law; while manipulated and erroneous data deviate from the law. Several factors may complicate our analysis. First, since an NPO may not cheat all the time (serial cheaters might be found out more quickly), 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 subtle types of cheating: such as rounding up numbers or petty manipulations which only affect 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 this in mind, we theoretically address the cost, if any, of 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).

Finally, 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.

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. In Appendix 8.6, we test this implication by excluding NPOs in the bottom 10%, 25%, and 50% of spending on governance. Another implication is that poor bookkeeping could lead to fewer data points being collected, so that including these NPOs would bias our results. We test this implication in Appendix 8.6 by alternating the threshold of the number of non-zero financial items used in our analysis from 65 to 115.

2.4.   An illustration of Benford's Law

Figure 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.[16] When each charity is considered, we fail to reject the null hypothesis that the observed distribution of the first digits of all the numbers follows the theoretical distribution for 75% of the sample using the Kolmogorov – Smirnov test at the significance level of 5%. Panel B provides a representative distribution of this group. We consider these NPOs as "conforming" to the law, namely, we fail to find evidence, both statistically and graphically, of potential misreporting. In contrast, we reject the null hypothesis for 25% of the sample. We call these charities as "deviating" from the law, suggesting that their full financial accounts contain inaccuracies detectable by Benford's Law which indicates potential misreporting. 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, we remove the charity's name. Panel C shows a clear graphical deviation from the Benford distribution, suggesting potential data manipulation.

---

[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.

Figure 1. The UK Third Sector Research Data and the Benford distribution
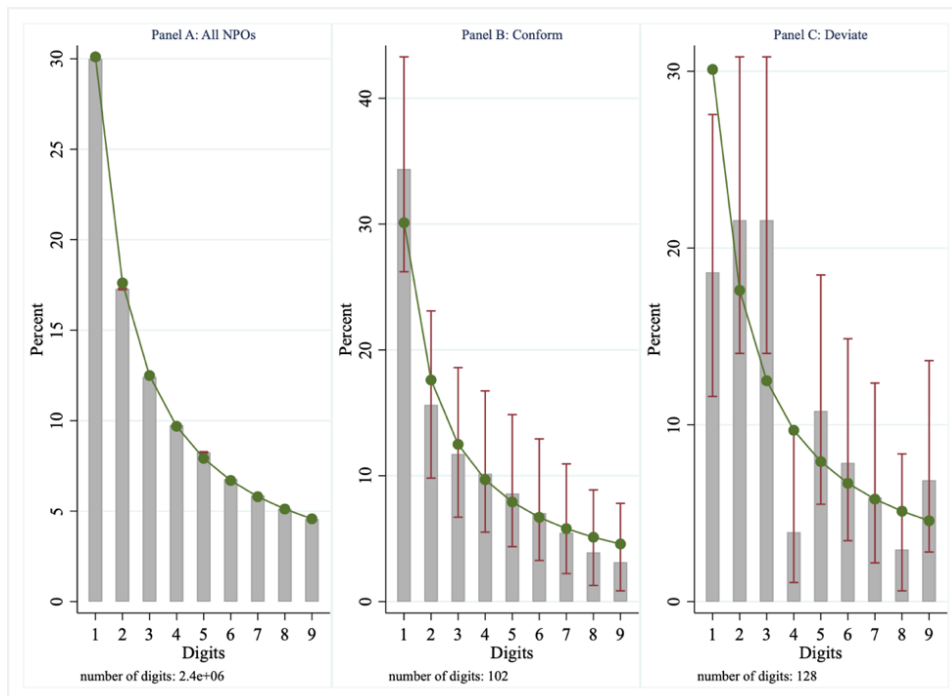
Note: Lines represent the theoretical distribution in Table 1. Bars represent the observed distributions in three samples. Capped spikes represent confidence intervals at the 95% significance level. In Panel A, we aggregate all the numbers in all financial accounts in all years provided by all the NPOs. 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. P-values used in the hypothesis testing are the exact p-values approximated by sampling from the null distribution (Monte Carlo simulation at 10,000 replication).

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 follows Goldman and Slezak's (2006) where the agent may take a hidden action, which affects the (actual) terminal value of the project and misreport the intermediate value (such as financial records of the organisation) to an uninformed donor to gain a higher payoff. In our context, the payoff can be either periodic grant disbursements subject to satisfactory performance of the agent or future grants that use the report as part of the grant 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.[18] We discuss key features of our model below.

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 NPO is assumed to be 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, the amount of unobservable action $a \geq 0$ (such as the level of dedication or effort). Second, the extent of misreporting (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

---

[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%.

[18] Human errors consist of unintentional coding errors or mistakes when inputting the numbers. Since our measures of misreporting rely on the very first digit of the numbers, it is unlikely that human errors would affect our measures of misreporting. Coding errors or rounding off numbers are more likely to affect the last few digits than the first digit. For completeness, however, we provide in the Online Appendix an extension of the model when we allow the level of governance spending to affect the organisation-specific human errors.

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 coming up with creative ways to get around monitoring requirements and (ii) the intrinsic aversion to lying (Hurkens & Kartik, 2009; Gneezy, 2005). Second, a higher level of governance spending g represents a higher governance quality in the organisation, leading to a higher disutility caused by the act of getting around the governance hurdles to manipulate the report. By governance quality, we follow Beyer et al.'s (2014) and Thakor's (2015) interpretation in which the oversight mechanism is fixed and required by the donor for pressurising the agent to conform to being accountable and aligning their interests more closely. For example, donors condition funding on the composition of the oversight board/committee; or the NPO's accounting division and choice of external auditors. In practical terms, governance spending involves governance-related 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 misreporting $b^e \geq 0$. The linear functionality is for tractability and captures two notions. First, the reputation loss $\psi^c(b)$ increases with misreporting, regardless of the prior belief $b^e \geq 0$. Since the NPO only suffers any reputation concern if the principal finds out, parameter c also captures the probability that the principal detects the deviation. The higher the probability of being found out, the higher the 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 parameter c 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 how 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 support 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 in the empirics.

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 η captures 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. 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, the 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 $t = 1$ (for example, the 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 summarises the model's timeline.

Figure 2. Timeline of the three stages.

	$t = 0$	$t = 1$	$t = 2$

- Donor offers a grant, cares about the terminal value V_2 net payment W
- NPO chooses action a and error b, faces non-profit constraint

- Firm-specific uncertainty η is observed
- $\theta^T = \rho a + \eta$ is the true state
- NPO issues $\theta = \theta^T + b$
- NPO receives grant W

- Market uncertainty ε is realised
- $V_2 = \rho a + \eta + \varepsilon - \lambda b$
- Donors receives payoffs $V - W$

17
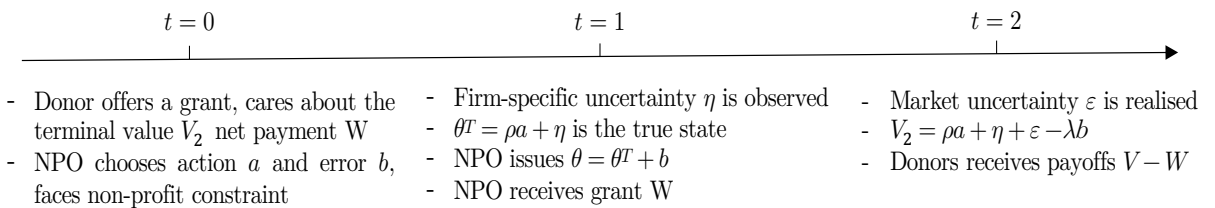
3.3. Payoffs

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 exists hidden action and information. The donor needs an indication of the project's terminal value to pay the 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,

---

[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.

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.

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 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 of the grants by solving:

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

subject to the incentive compatibility $\{a^*, b^*\} = \arg\max_{a,b} CE$ 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):

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

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 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, $g > \delta\lambda/\rho^2$, 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 (i) and (ii) 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.

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 productivity (ρ) of the NPO is sufficiently high or the disutility (δ) is relatively small 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. $\qquad \left.\frac{\partial b^*}{\partial a^*}\right|_{\partial g} > 0$ if and only if $g < \tilde{g}$

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

Proof: See Appendix 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, an 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 $w_1^* > 0$, there exists a 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 8.3.

Proposition 2 suggests that higher spending on governance activities (a tighter oversight mechanism) does not guarantee reports being more credible. An 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, an 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. First, 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 a higher possibility of being found out or they place higher importance on 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 $w_1^* > 0$, 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 misreporting, rather the effect is through 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 the risk. The NPO would now have less incentive to manipulate the report due to the lower value-based incentive. The 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 effort to improve the true state of the project and manipulate the report on which the contract is based.

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 the level of optimal misreporting would depend on both the charitable effort and the oversight mechanism (cost of misreporting). We test the predictions below.

4. The econometric methodology

4.1. Empirical specifications

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 $I_i = a_i \times g_i$ 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 vectors of control variables and their 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 $g = 0 < \tilde{g} =$

$-\gamma_3/\gamma_1$, Proposition 1 suggests that $\gamma_1 = \partial b^*/\partial a^* > 0$. Likewise, when the NPO chooses zero effort, $a^* = 0 < \tilde{a} = \gamma_3/\gamma_2$, Proposition 2 suggests (Appendix 8.3) that $\gamma_2 = \partial b^*/\partial g > 0$. If g is sufficiently high such that: $\partial b^*/\partial a^* = \gamma_1 + \gamma_3 g < 0$, 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 we take the averages of the respective financial figures over the years, to correspond to the pooled annual financial statements constructed for the dependent variable. To proxy for each NPO's choice of the optimal charitable action we use the ratio of spending on charitable activities to the NPO's total income (*Charitable Spending)*. 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. This has long been used as the standard measure for the effectiveness of a charity (namely, the programme ratio as reviewed in Hofmann and McSwain, 2013).

For *Governance Spending*, g_i, we use the proportion of the total income spent on governance activities. This variable captures the governance-related overhead costs as modelled in Section 3. As reported in Table A1, there are two categories underneath the broad item *Total Governance Costs*: (i) *Costs of accounting and audit fees*, and (ii) *Costs of administration*. According to the reporting practice recommended by the Charity Commission (SOPR, 2005:p.31) these costs include: the cost of accounting activities and the fees of internal and external audits; legal advice for trustees; costs of holding statutory and board meetings; and governance-related administrative costs, such as, preparing statutory accounts, some employee costs where staff are providing administrative support to trustees, or costs of materials in supporting the governance activities. To capture the quality of the oversight and governance mechanism, we would ideally want to use the size of the governing board, or the professional certificates or years of experience held by the governing board. In the absence of such variables, we argue our aggregated spending variable acts as a good proxy for the quality of the oversight mechanism. As explained in the theoretical model, our aggregated variable reduces the level of mistakes in two ways: first, the higher the aggregated governance costs the higher the disutility of misreporting due to a higher probability of getting caught from the increased level of auditing and administration; second a higher governance spending would improve the book-keeping procedures that can ensure accurate records and reduce human errors. We experiment with using each component of the aggregated costs in Table OA1 (Online Appendix) and obtain 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 up to 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 instruction or may simply 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 strength of the non-profit's philanthropy arm and concern about their reputation. The consequence is the organisation becomes more open and 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 in our empirical strategy below.

We include six additional 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*)[22], (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 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 variables are expected to be associated with misreporting. Not having reports audited is a popular potential determinant of accounting fraud; while 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

---

[22] We also replace the dummy variable *Zero fundraising* with the actual spending on fundraising and publicity activities (average over the years) and re-do our analysis. The qualitative results (insignificant estimates) remain.

implausible that a non-profit could incur exactly zero expense in fundraising, hence reporting zero fundraising should infer some reporting inaccuracy. 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. We expect that as these charities have 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.[23]

To further test the impact of fundraising pressure and reputation concern outlined in Corollary 3 (part i), we include the variable *Income from donations/grants*, constructed as the ratio of the income from charitable activities and voluntary sources over total income, to capture the intensity of the reliance on these income sources.

Finally, we include the number of non-zero financial entries (*Number of non-zeros*) and the number of annual reports (*Number of yearly reports*) to control for the size of the digit pool used to create the accuracy measures. Controlling for NPO's size and the number of yearly reports, more non-zero financial items being reported could be an indication of the NPO's openness and transparency. For that reason, we expect *Number of non-zeros* to be associated with a higher level of accuracy.[24] For two NPOs with the same *Number of non-zeros,* the NPO with fewer *yearly reports* could either have fewer activities or a less complex operation or be strategically withholding information by recording more zeros in their annual report. We expect the *Number of yearly reports* to be associated with a higher level of inaccuracy.

---

[23] A pitfall of using binary variables is the loss of information. However, 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 long period. Binary variables could retain the information of, for example, whether the NPO experiences losses. Such measures are less prone to measurement error. 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.

[24] Our main results remain similar even when we exclude these two control variables.

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 the *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 mis-measured, with this measurement error increasing the *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 attenuation bias. Our estimates would become closer to zero than the unbiased parameters, but the signs should remain. We aim to address this 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 further by employing two IV strategies, namely, the traditional method and Lewbel's heteroscedasticity IV esstimator.

The traditional IV approach requires valid instruments that satisfy two criteria: being strongly correlated with *Charitable Spending* and the *Interaction term* (strong identification) and be 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 free from possible misreporting. Based on the data available, we propose two instruments: the NPO's number of official 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 staffs is easily either observed or can be 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.[25] 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 the *Interaction term* could be endogenous due to the *Charitable Spending* component. 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.

The second IV approach is proposed by Lewbel (2012). The estimator exploits heteroscedasticity and higher moment conditions to construct internal instruments from the model's data without the need for any external source of variation. 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. 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 and Hou, 2013; Loy et al., 2016; Caliendo et al., 2017), and

---

[25] 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.

supplement the set of internal instruments with our instruments to improve the efficiency of the estimator. We describe the estimator's intuition in Appendix 8.5.

Finally, the recorded spending on accounting and auditing services could also be subject to both measurement error 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 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 size, volunteers, and whether the reports are audited. Another concern is that *Governance Spending* could be endogenous, even though it is a sectoral norm that governance costs are relatively fixed and required by external bodies to fulfil their legal duties (see Hyndman and McConville, 2016, and Thakor, 2015 for a theoretical account). For completeness, we experiment with 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.

5. Empirical results

5.1. Descriptive statistics

Table 2 presents descriptive statistics and Table 3 compares 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 of deviation from the Benford distribution, as measured by MAD, χ^2, and KS statistics. Over the period of 2007 – 2015, on average, UK NPOs spend 76% of their annual income on charitable activities, and 4% on governance activities. The maximum values for the proportions of the total income spent on the two activities are large (nearly 800% and 100% of the total income, respectively). It is possible for NPOs to spend eight times their income on these activities if they have access to endowments. Of the sampled NPOs, 16% receive endowment funds. There is a clear difference in charitable spending between conforming and deviating NPOs. Deviating NPOs appear to spend more on charitable spending, be older, submit more annual reports, but receive less government or restricted income, and record more losses from investment. The figures for governance spending, size (total assets) 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, supporting our intuition that these variables should not directly correlate with the misreporting behaviour of the NPOs.

Figure 3 plots the histograms of the three variables of interest: the *MAD statistic*, *Governance Spending* and *Charitable Spending*. Unsurprisingly, 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 statistic*, resembling a normal distribution.[26] We report other descriptive statistics in Table A2.

Figure 3. Histograms of the measures of misreporting and key expenses

Panel A. Histograms of Governance Spending and Charitable Spending

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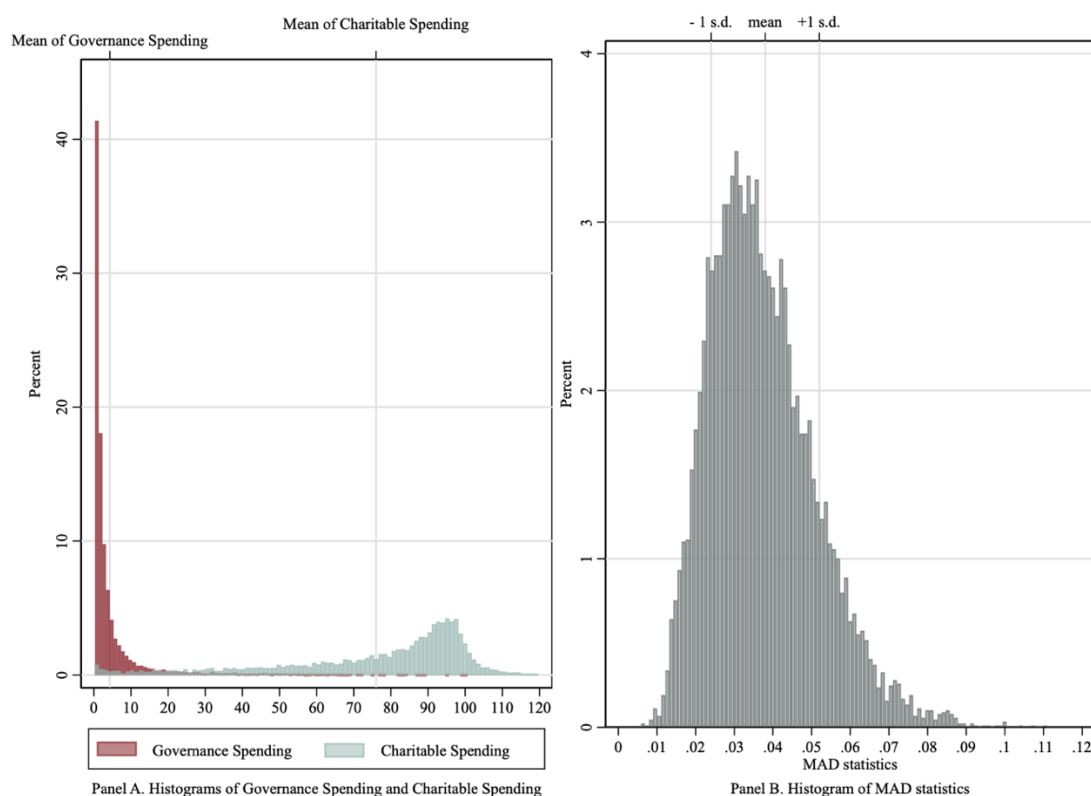
Note: Panel A plots histograms of *Governance Spending and Charitable Spending* as the percentages of the total income. Panel B plots the histogram of the *MAD statistic*. Number of bins is 100 for both panels. For presentation purposes, we exclude the top 1% of each variable from the histograms.

---

[26] Figure A2 in the Appendix 8 plots the statistics over five quantiles of key NPO characteristics such as charitable spending, governance spending, and size (total assets).

Table 2. 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
Headcounts	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		

Note: *** 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 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.

5.2. Regression results

Table 3 shows estimates from an OLS with full control variables, the 2SLS with our proposed instruments, and the Lewbel's (2012) estimator.[27] 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 test. If we fail to reject the null of statistical equivalence when treating the *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. In Table 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. Column 3 presents our preferred 2SLS specification with control variables. 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 traditional 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 A4 in the Appendix.

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 and dependent on the magnitudes (the thresholds) of the other expense. To determine the thresholds in our dataset for UK charities, we derive the following thresholds from Equation (21).

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

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

Figure 4 visualises the values of these thresholds. Calculations, using the OLS with control variables and the Lewbel's estimation (our preferred specification), suggest that $\tilde{a} \approx 70\%$ and

---

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

$\tilde{g} \approx 15\%$. That is, only organisations who exert relatively high effort in charitable activities (spending at least 70% of their total income), would have a lower measure of misreporting when they spend more on governance activities. Otherwise, higher governance spending need not translate into a higher level of reporting accuracy. Similarly, only when charities spend sufficiently on auditing and accounting activities (about 15% of total income), will charities performing well (spending more on charitable activities) be associated with a lower level of misreporting. To put these numbers into perspective, only 7% of the charities in our sample spend more than 15% of their total income on governance activities. Despite being suggestive, our threshold of 15% highlights the lack of resources spent on governance activities in the UK sector. The result supports the call for increased support from granting donors for overhead costs and accounting expenses, as documented by Singh (2015).

Figure 2. The predictive marginal effects on inaccuracy of financial reports
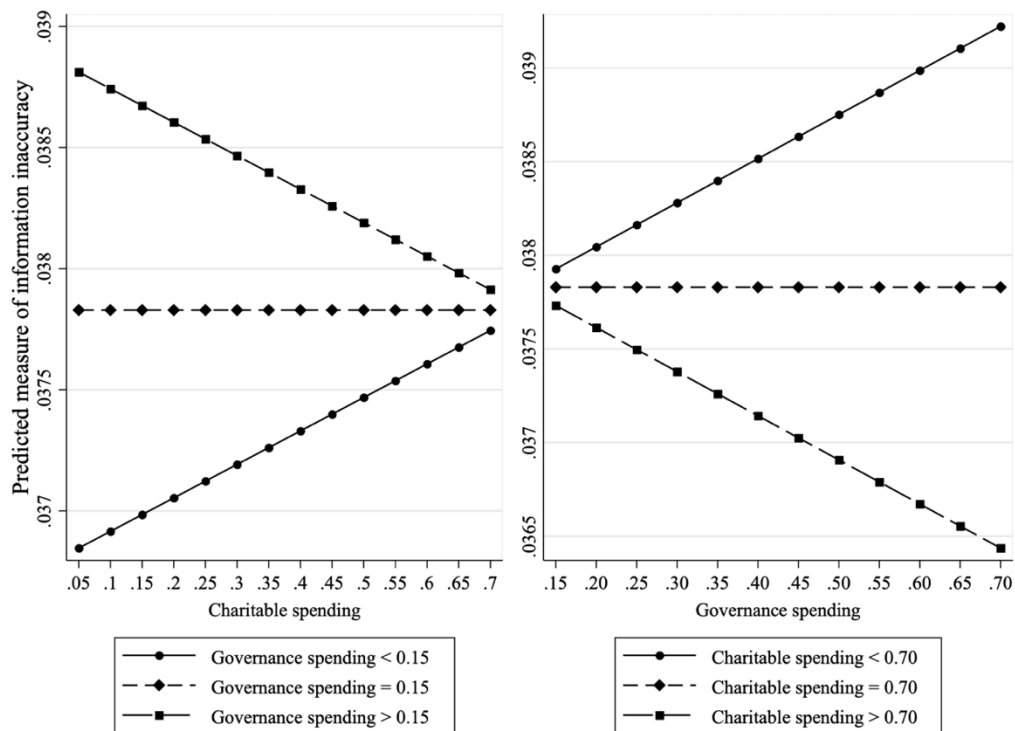
Note: Margin plots of the marginal effects obtained after our preferred Lewbel (2012) estimation. The left (right) panel shows marginal effect of charitable spending (governance spending) on the predicted measures of misreporting. Lines with cirles (squares, diamonds) represent the respective marginal 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 3 also provides interesting estimates for the other organisational characteristics. Contradictory to Krishnan et al. (2006), size and age are positively correlated with the amount of information irregularities. The result is not 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 with 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 income from the public or grants, however, report more accurately. This result is consistent with the explanation of reputation concern – those who would like to remain trustworthy are more likely to behave well. Having more non-zero financial items recorded, while 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 NPOs reporting truthfully. In Appendix 8.6.4, we find similar results when we control for the possibility that some NPOs report fewer non-zero transactions to withhold information.

Table 4 supports the statistical validity of our instruments. The Kleibergen-Paap Wald F-statistics 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 the C-statistic to test for the orthogonality of suspect instruments (see Hayashi, 2000, pp. 227-8). C-statistics provide additional information over Hansen J statistics. For a model with many instruments, a Hansen-Sargan test may have little power (Baum et al., 2003). In addition, 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. Both the Hansen J statistics and C-statistics in Table 4 fail to reject the null of overidentification, supporting the statistical validity of our 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 3. Main results for the determinants of misreporting for British charities

VARIABLES	*Dependent variable:* MAD statistic			
	(1) OLS	(2) OLS with controls	(3) 2SLS	(4) Lewbel's (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**
(% of total income)		(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

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. 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. The unit of observation is individual British NPO.

Table 4. 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 (p-value)	54.461	306.401
	(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	6.800
	(0.641)	(0.147)
(endogeneity) Wu-Husman test of endogeneity Chi-square (1) (p-value)	5.466	6.575
	(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)

Note: Kleibergen-Paap 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.

6.   Sensitivity Analyses

In this section, we present an exercise using within-NPO variations to compliment the between-NPO comparison above. Since the data span an eight-year period, we can separate the sample into two periods (2008 – 2011 and 2012 – 2015) while maintaining the threshold of non-zeros (100). We first split the sample into two four-year intervals and re-calculate the measures of misreporting and the explanatory variables as described in Section 4.1 for each. 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).

Despite losing observations, the within-NPO exercise has three advantages over the between-NPO comparison. First, the within-NPO analysis allows us to control for both time-variant and time-invariant characteristics of the NPOs. Since the annual reports were collected in waves, the NPOs could have organisation-specific reasons to misreport the financial statements, such as the inherent tendency to manipulate the financial reports or to make human errors (accounting competence). By focusing on the dynamics of the *Charitable Spending* and *Governance Spending* variables over the two periods we mitigate concerns of the measurement error bias that could result from the tendency of the NPOs to manipulate these variables in both periods. Second, by controlling for time fixed effects, we mitigate concerns that changes in macroeconomic conditions in the UK and the charity sector have driven our results. Finally, by having two different intervals, we demonstrate that our results are not driven by the aggregation of the data.

Table 5 presents our results. Controlling for within-NPO time-invariant and time-varying NPO characteristics (omitted to save space), we obtain the same thresholds as in our main analysis. Columns (2) and (3) 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.

We perform four additional sensitivity checks in Appendix 8.6 and another two in the Online Appendix. While the precise magnitudes of the estimates 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.

Table 5. Fixed effect models for misreporting over two periods

VARIABLES	Dependent variable: *MAD statistic*		
	(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 4.1. Control variables are listed in Table 2 or Section 3.1

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 measure of financial misreporting. 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 rising 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. A higher requirement of governance spending and stricter oversight mechanism would be effective in improving the accuracy of financial reporting only when the NPOs are already incentivised to exert more effort (for example, a larger share of income spent on charitable activities). Our preferred estimates suggest that spending at least 15% of the total income on governance activities would help better performing charities provide reports that are more accurate. Unless funders and the public consider support for these governance 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 packages to overhead activities. We hope to open further research and discussions on these issues.

8. Appendix

Table 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
Incoming resources (Total)*
Charitable activities – Government Sector
Charitable activities – Central Government
Charitable activities – Local Government
Charitable activities – Regional Government
Charitable activities – Town and Parish Councils
Charitable activities – NHS Trusts
Charitable activities – European Government
Charitable activities – International Government
   Agency
Charitable activities – Foreign Governments
Charitable activities – Public Corporations
Charitable activities – Universities
Charitable activities – Devolved Governments
Charitable activities – Business Sector
Charitable activities – Nonprofit sector
Charitable activities – General public
Charitable activities – Government Sector

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

Other income
Restricted income
Unrestricted income
Other financial values
Gains/Losses on pension funds
Reserves
Staff costs
Other staff costs
Pension costs
Social security costs

Wages and salaries
Support costs
Irrecoverable VAT
Tangible fixed assets
Additions
Net book value – beginning
Disposals
Net book value – end

Note: * 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.

Table A2. 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	0.062	0.014	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 (£'000)	77.294	361.013	0	1.517e+04

Note: The sample is restricted to 10322 British charities whose number non-zero financial items in their reports is at least 100.

Figure A1. Distribution of skewness and mean-to-median ratio of the aggregated data

Panel A: Positive skewness of the aggregated data for each NPO

Panel B: Mean > Median for the aggregated data of each NPO
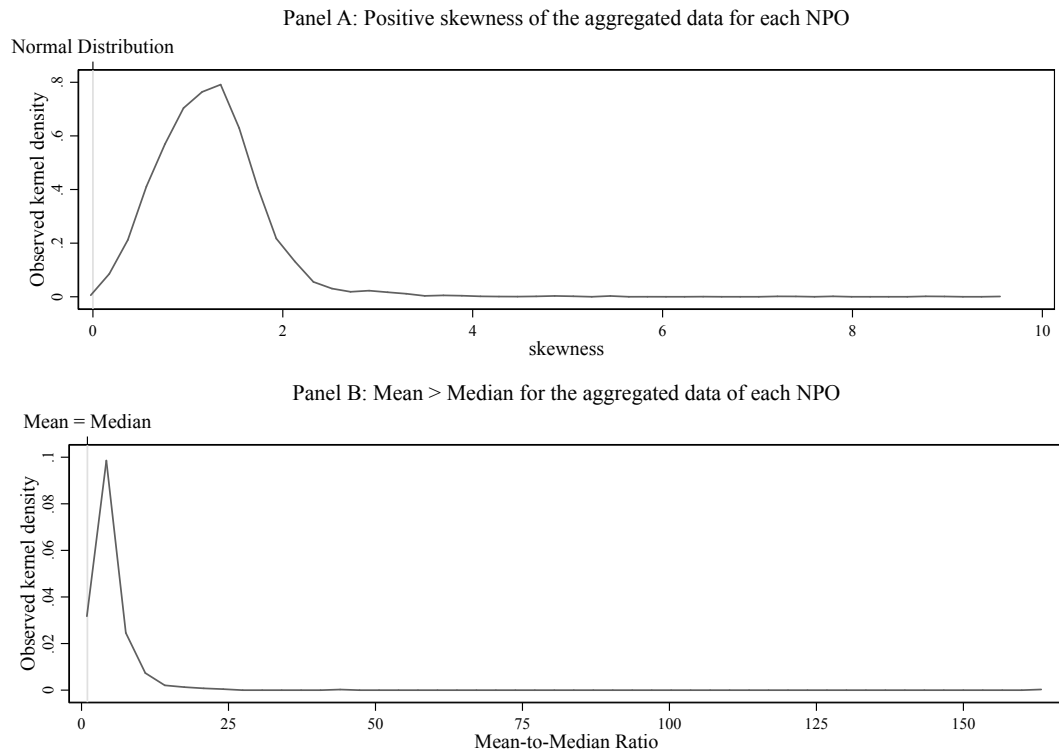
Note: The aggregated data are conducted by aggregating the yearly financial data of each charity. The figure plots the distributions of skewness values and mean-to-median ratios for the 10,322 charities that have at least 100 non-zero financial items in their aggregated data. The two panels clearly demonstrate that the aggregated data for each charity have (1) positively skewed distribution (Panel A) and (2) mean larger than median (Panel B). As discussed in Durtschi et al. (2004), the data also should conform with Benford's Law.

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 $\omega = N(\mu, d^2)$ 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:

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

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

(A1) and (A2) imply that:

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

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

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

8.2. Proof of Proposition 1

We first notice that:

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

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:

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

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

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

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

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

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 A2 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), $\frac{\partial b^*}{\partial a^*}\bigg| \partial g = \dfrac{\frac{\partial b^*}{\partial g}}{\frac{\partial a^*}{\partial g}} < 0$ if and only if $g \geq g_2$ and $\frac{\partial b^*}{\partial a^*}\bigg| \partial g > 0$ if and

only if $g < g_2$. Set $g_2 = \tilde{g}$ and the proof completes.

Figure A2. How sign $\frac{\partial b^*}{\partial g}$ and sign T(g) behave when g varies in $(0, +\infty)$.

sign $\frac{\partial b^*}{\partial g}$ = sign $T(g) > 0$ if $g < \tilde{g}$

$T(g)$

g_{max}

g_1

0

$g_2 = \tilde{g}$

g

sign $\frac{\partial b^*}{\partial g}$ = sign $T(g) < 0$ if $g > \tilde{g}$
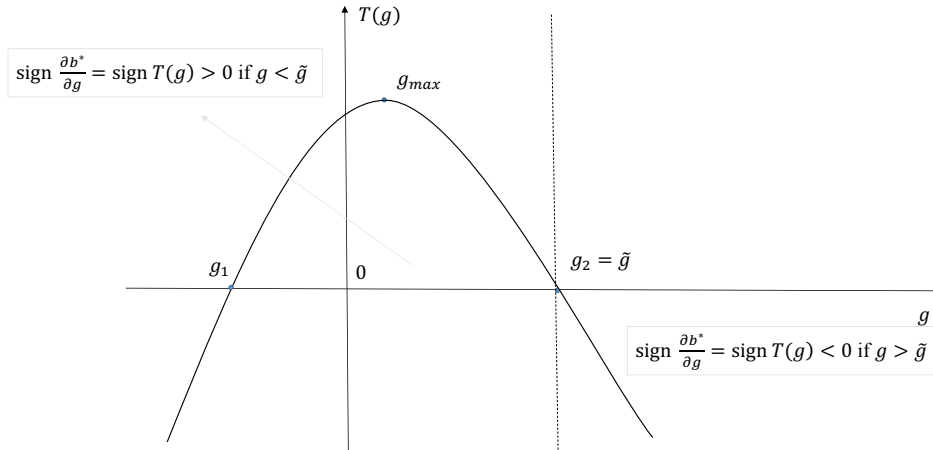
Source: Authors' own illustration.

8.3. Proof of Proposition 2

Proof: As $\frac{\partial a^*}{\partial g} > 0 \ \forall g \geq 0$, $g \geq \tilde{g}$ 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 A2. Figure A2 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.

8.4. Additional summary statistics

Figure A3. Measures of misreporting (MAD) by quantiles of NPO characteristics
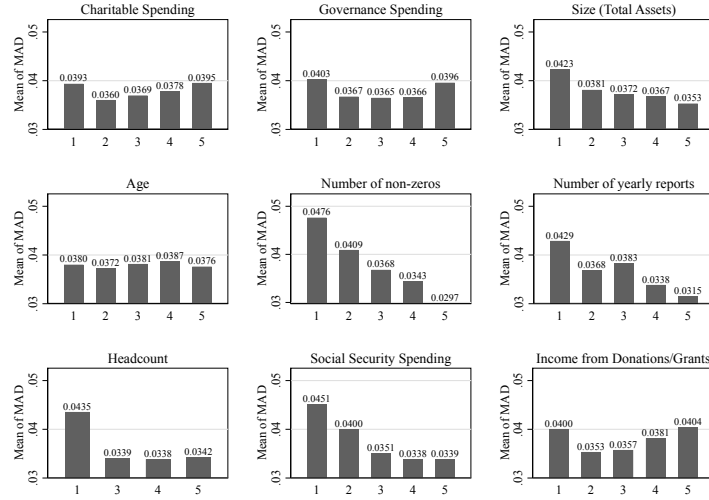
Notes: The *MAD statistic* is plotted over five quantiles of nine NPO characteristics.

8.5. Lewbel's (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 $Z_{ij} = (X_i - \bar{X}_i)\varepsilon_{ji}$ 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 (optional).

8.6. Robustness checks

8.6.1. Sensitivity to the measurement of reporting inaccuracy

We replace the *MAD statistic* by three 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 A4 reports the estimates from OLS with full controls, our preferred 2SLS, 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.

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. 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.

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 A5 summarises 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 A4. Robustness to varying the cut-offs of non-zero financial points
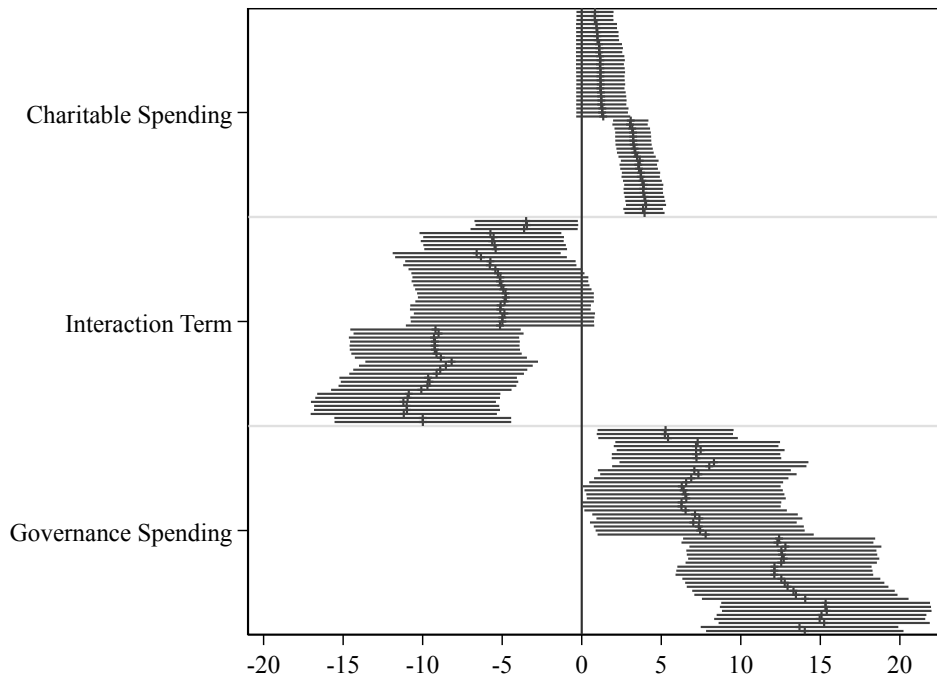
Note: The estimates of the three variables of interest are in a descending order of the number of non-zeros used from 65 to 115. For example, the top estimate is for Charitable Spending at the cut-off of 65. Confidence level is at 95%. The graph shows a clear robustness to our main results when varying the cut-offs.

Figure A5. Robustness to excluding various ranges of outliers
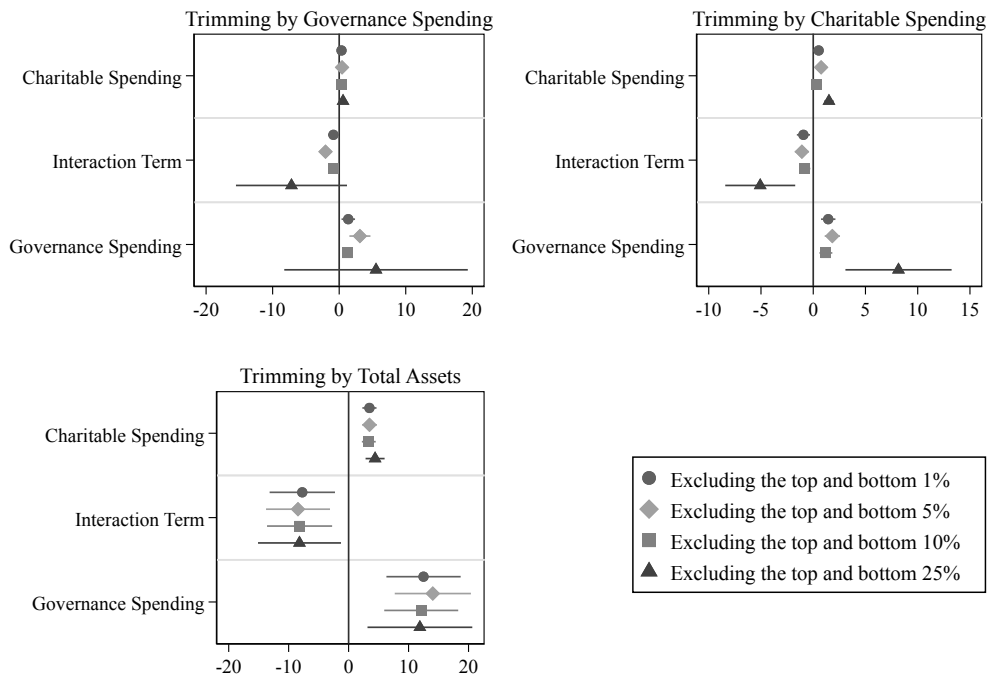
Note: The checks use OLS with full controls as in the main analysis.

8.6.4. Controlling for potential informational non-disclosure

Since we exclude NPOs with fewer than 100 non-zero financial items, sample selection could be an issue. The excluded NPOs could either have operated in more straightforward activities which generate no significant transactions or have strategically withheld information by recording some significant items as zeros. Although our main results are not sensitive to the threshold of 100, we use the Heckman sample correction to check our results' 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 the selection of NPOs that 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 the predicted individual probabilities in the first stage as additional explanatory variables. Table A3 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 statistic*, 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 A3. 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 Spending	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)

Note. *** p<0.01, ** p<0.05, * p<0.1, robust standard errors are in parentheses. First-stage estimates probit of T (= 1 if included in the digital analysis as having at least 100 non-zeros, 0 otherwise). Second stage follows Heckman's (1979). Chi-square (1) statistics are for Wald test of independence (rho) of two stages.

Table A4. Replications using other indices

VARIABLES	χ^2 OLS	χ^2 2SLS	χ^2 Lewbel's	KS statistics OLS	KS statistics 2SLS	KS statistics Lewbel's	Deviate (1 = Deviate, 0 = Conform) OLS	Deviate (1 = Deviate, 0 = Conform) 2SLS	Deviate (1 = Deviate, 0 = Conform) 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 Spending	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)

Note: 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.

51

8.7.   Additional Tables

Table A5 compliments Table 4 in the main text for estimates of control variables for various estimators. We experiment with treating *Governance Spending* and the *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 estimations, 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 A5. 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 Spending	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	-	-
Treat the Interaction term as exogenous?	No	Yes	No	Yes	Yes	No
Treat Governance spending as exogenous?	Yes	Yes	Yes	Yes	No	No
K-P F-stat for weak identification	16.77	15.95	10.49	9.047	110.90	20.67
Hansen J statistics/C-statistics (p-value)	-	0.45 [0.50]	-	0.00 [0.98]	3.40 [0.17]	16.26 [0.00]
Endogeneity test for Interaction term	0.45 [0.50]	-	4.78 [0.03]	-	-	2.52 [0.12]
Endogeneity test for Governance Spending	-	-	-	-	0.002 [0.96]	1.83 [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. Tests of endogeneity: Ho: the specified endogenous can be treated as exogenous (Durbin-Wu-Hausman).

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.

Aebischer, P. (2012, January 19). Philanthropy: The price of charity. *Nature*, Vol. 481, p. 260. https://doi.org/10.1038/481260a

Aldashev, G., & Navarra, C. (2018). Development NGOs: Basic facts. *Annals of Public and Cooperative Economics*, *89*(1), 125-155.

Aldashev, G., Marini, M., & Verdier, T. (2015). Governance of non-profit and non-governmental organizations–within-and between-organization 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.

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.

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.

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

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

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *The American Economic Review*, *96*(5), 1652-1678.

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

Beyer, A., Cohen, 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.

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.

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.

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.

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

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.

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.

Cho, W. K., & Gaines, B. J. (2007). Breaking the (Benford) law: Statistical fraud detection in campaign finance. *The American Statistician*, 61(3), 218-223.

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

Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, *50*(2), 344-401.

Durtschi, C., & 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., and Pacini, C. (2004), The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensic Accounting*, 5, 17–34.

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.

Exley, C. L. (2019). Using Charity Performance Metrics as an Excuse Not to Give. *Management Science*, mnsc.2018.3268. https://doi.org/10.1287/mnsc.2018.3268

Fang, H., & Gong, Q. (2017). Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked. *The American Economic Review*, *107*(2), 562-591.

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.

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.

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.

Gneezy, U., Keenan, E. A., & Gneezy, A. (2014). Avoiding overhead aversion in charity. *Science*, *346*(6209), 632–635. https://doi.org/10.1126/science.1253932

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.

Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, *31*(1), 405-440.

Hermalin, B. E., & Weisbach, M. S. (2012). Information Disclosure and Corporate Governance. *The Journal of Finance*, *67*(1), 195-233.

Hill, T.P (1995). The first digital phenomenon. *American Scientist.* 86(4):368-363.

Hofmann, 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.

Hyndman, N., & McConville, D. (2016). Transparency in reporting on charities' efficiency: A framework for analysis. *Nonprofit and Voluntary Sector Quarterly*, *45*(4), 844-865.

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.

Judge, G., & Schechter, L. (2009). Detecting problems in survey data using Benford's Law. *Journal of Human Resources*, 44(1), 1-24.

Keating, E. K., & Frumkin, P. (2003). Reengineering nonprofit financial accountability: Toward a more reliable foundation for regulation. *Public Administration Review*, *63*(1), 3-15.

Keating, K., Parsons, M., & Roberts, A. (2008). Misreporting fundraising: How do nonprofit organisations account for telemarketing campaigns? *The Accounting Review*, *83*(2), 417-446.

Kossovsky, A. E. (2015). *Benford's Law*. Singapore: World Scientific Publishing.

Krishnan, R., & Yetman, M. H. (2011). Institutional drivers of reporting decisions in nonprofit hospitals. *Journal of Accounting Research*, *49*(4), 1001-1039.

Krishnan, R., Yetman, M. H., & Yetman, R. J. (2006). Expenses misreporting in non-profit organisations. *The Accounting Review*, *81*(2), 399-420.

Lee, D. S., & Lemieuxa, T. (2010). Regression discontinuity designs in economics. *Journal of Economic L iterature*, *48*(2), 281-355.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, *30*(1), 67-80.

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 & Organization*, *123*, 184-192.

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.

Miller, S. J. (Ed.). (2015). *Benford's Law: Theory and Applications*. Princeton University Press.

Morrow, L. (2014). *Benford's Law, families of distributions and a test basis. CEP Discussion Papers*, CEPDP1291. Centre for Economic Performance, LSE.

Nigrini, M. J. (1996). A taxpayer compliance application of Benford's law. *The Journal of the American Taxation Association*, *18*(1), 72.

Nigrini, M. J. (2011). *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations* (Vol. 558). John Wiley & Sons.

Nigrini, M. J. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing and Fraud Detection.* Hoboken, New Jersey.

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.

Planck, M. (1901). On the law of the energy distribution in the normal spectrum. *Ann. Phys*, *4*(553), 90.

Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics, 85*(4), 777–792.

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.

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.

Senchaudhuri, P., Mehta, C. R., & Patel, N. R. (1995). Estimating exact p values by the method of control variates or Monte Carlo rescue. *Journal of the American Statistical Association*, *90*(430), 640-648.

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.

SOPR. (2005). Accounting And Reporting By Charities: Statement Of Recommended Practice. Retrieved January 8, 2020, from 2005 website: https://www.gov.uk/government/publications/charities-sorp-2005

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

Thakor, A. V. (2015). Strategic information disclosure when there is fundamental disagreement. *Journal of Financial Intermediation*, *24*(2), 131-153.

Trussel, J. (2003). Assessing potential accounting manipulation: The financial characteristics of charitable organizations with higher than expected program-spending ratios. *Nonprofit and Voluntary Sector Quarterly*, *32*(4), 616-634.

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.

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*.

Wedig, G. J. (1994). Risk leverage donations and dividends-in-kind: A theory of nonprofit financial behavior. *International Review of Economics and Statistics,* 3, 257–278.

Woodwell, H.W., & Bartczak, L. (2008). *Is Grantmaking Getting Smart? A National Study of Philanthropic Practice.* Washington, D.C.: Grantmakers for Effective Organisations.

Yetman, M. H., & Yetman, R. J. (2012). The effects of governance on the accuracy of charitable expenses reported by nonprofit organizations. *Contemporary Accounting Research*, *29*(3), 738-767.

Zitzewitz, E. (2012). Forensic economics. *Journal of Economic Literature*, 50(3), 731-769.

Does transparency come at the cost of charitable services?
Evidence from investigating British charities

Canh Thien Dang

Department of Economics

London School of Economics, UK

Trudy Owens

School of Economics

University of Nottingham, UK

1. Extension of the theoretical model

We extend the theoretical model in the main analysis by allowing the level of unintentional human errors to be affected by the level of governance spending. The argument here is, despite not being established in the empirical literature, it could be argued that higher spending on governance and accounting activities could reduce the probability of human errors (such as coding errors or mistakes when inputting the numbers). Since this type of error occurs in specific organisations, we assume that an increased level of governance spending would reduce the variance of the organisation-specific uncertainty η of the intermediate value θ. That is, the performance measure of the organisation is less likely to be misreported due to random factors. We obtain a similar result for the existence of the thresholds.

To prove the existence of the thresholds, we need to make the following assumption.

Assumption OA1: For simplicity, we assume that the variance linearly decreases in the level of governance spending $\eta(g) \sim N(0, \frac{\sigma_\eta^2}{1+g})$ and the uncertainty is maximised when the NPO does not spend on governance spending $\eta(0) \sim N(0, \sigma_\eta^2)$, as in the main analysis.

In contrast to the main analysis, we are not able to derive closed equilibria for the statics of interest. However, we show that the main theoretical predictions of the existence of the thresholds remain.

Indeed, in replacing the new organisation-specific uncertainty, the new maximisation problem becomes:

$$\max_{w_1} \frac{\rho^2}{\delta} w_1 - \lambda \left[\frac{(w_1 - c)}{g} + \tau \right] - \frac{\rho^2 w_1^2}{2\delta} - \frac{(w_1 - c)^2}{2g} - c \left[\frac{w_1 - c}{g} - b^e \right] - \frac{r}{2} \frac{\sigma_\eta^2}{1+g} w_1^2 \quad (OA1)$$

Solving the equation (OA1) for w_1, the new value-based incentive becomes:

$$w_1^* = \frac{\dfrac{\rho^2}{\delta} - \dfrac{\lambda}{g}}{\dfrac{\rho^2}{\delta} + \dfrac{1}{g} + \dfrac{r\sigma_\eta^2}{1+g}} \tag{OA2}$$

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

$$b^* = \frac{\beta - \dfrac{\lambda}{g}}{g\beta + \dfrac{\Delta g}{1+g} + 1} - \frac{c}{g} + \tau \tag{OA3}$$

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

Proposition OA1: When $w_1^* > 0$, there exists a fixed threshold of the *governance spending* $\tilde{g} > 0$ such that:

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

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

Proof: The intuition is similar to the main model when the level of governance spending enters the optimal level of misreporting non-monotonically. For a formal proof, we first rewrite:

$$\left.\frac{\partial b^*}{\partial a^*}\right|_{\partial g} = \frac{\dfrac{\partial b^*}{\partial g}}{\dfrac{\partial a^*}{\partial g}} \tag{OA4}$$

We rewrite a^* as:

$$a^* = \frac{\rho}{\delta} w_1 = \frac{\rho}{\delta}\left(\frac{\frac{\rho^2}{\delta} - \frac{\lambda}{g}}{\frac{\rho^2}{\delta} + \frac{1}{g} + \frac{r\sigma_\eta^2}{1+g}} \right) = \frac{\rho}{\delta}\left(\frac{(1+\lambda)\frac{\rho^2}{\delta} + \frac{\lambda r\sigma_\eta^2}{1+g}}{\frac{\rho^2}{\delta} + \frac{1}{g} + \frac{r\sigma_\eta^2}{1+g}} - \lambda \right)$$

$$= \frac{\rho}{\delta}\left(\frac{\frac{(1+g)(1+\lambda)\rho^2}{\delta} + \lambda r\sigma_\eta^2}{\frac{\rho^2}{\delta} + 1 + r\sigma_\eta^2 + g\left(\frac{\rho^2}{\delta}+1\right)} - \lambda \right)$$

$$= \frac{\rho}{\delta}\left(\frac{\frac{(1+\lambda)\rho^2}{\delta} + \lambda r\sigma_\eta^2 + \frac{g(1+\lambda)\rho^2}{\delta}}{\frac{\rho^2}{\delta} + 1 + r\sigma_\eta^2 + g\left(\frac{\rho^2}{\delta}+1\right)} - \lambda \right)$$

$$= \frac{\rho}{\delta}\left(\frac{\frac{(1+\lambda)\rho^2}{\delta} + \lambda r\sigma_\eta^2 + \frac{g(1+\lambda)\rho^2}{\delta}}{\frac{\rho^2}{\delta} + 1 + r\sigma_\eta^2 + g\left(\frac{\rho^2}{\delta}+1\right)} - \frac{(1+\lambda)\rho^2}{\rho^2+\delta} + \frac{(1+\lambda)\rho^2}{\rho^2+\delta} - \lambda \right)$$

$$= \frac{\rho}{\delta}\left(\frac{\frac{(1+\lambda)\rho^2}{\delta} - \frac{(1+\lambda)\rho^2}{\rho^2+\delta}(\frac{\rho^2}{\delta} + 1 + r\sigma_\eta^2)}{\frac{\rho^2}{\delta} + 1 + r\sigma_\eta^2 + g\left(\frac{\rho^2}{\delta}+1\right)} + \frac{(1+\lambda)\rho^2}{\rho^2+\delta} - \lambda \right)$$

Since $\frac{(1+\lambda)\rho^2}{\delta} < \frac{(1+\lambda)\rho^2}{\rho^2+\delta}(\frac{\rho^2}{\delta} + 1 + r\sigma_\eta^2)$ and $\frac{1}{\frac{\rho^2}{\delta}+1+r\sigma_\eta^2+g\left(\frac{\rho^2}{\delta}+1\right)}$ decreases with g, again we have

$$\frac{\partial a^*}{\partial g} > 0, \qquad \forall g \geq 0 \tag{OA5}$$

From Equation (OA3):

$$b^* = \frac{\beta - \frac{\lambda}{g}}{g\beta + \frac{g\Delta}{1+g} + 1} - \frac{c}{g} + \tau \tag{OA6}$$

Differentiating b^* with respect to g we have:

$$\frac{\partial b^*}{\partial g} = \left(\frac{\beta - \frac{\lambda}{g}}{g\beta + \frac{g\Delta}{1+g} + 1} \right)' + \frac{c}{g^2}$$

$$= \frac{\frac{-\lambda\Delta}{1+g} + \lambda(\Delta+1) - \Delta\beta g^2 - \frac{\Delta g^2}{(1+g)^2} + \frac{\lambda g}{(1+g)^2} + c\left(g\beta + \frac{g\Delta}{1+g} + 1\right)^2}{g^2\left(g\beta + \frac{g\Delta}{1+g} + 1\right)^2}$$

$$= \frac{T(g)}{g^2\left(g\beta + \frac{g\Delta}{1+g} + 1\right)^2}$$

$$\rightarrow \text{sign} \ \frac{\partial b^*}{\partial g} = \text{sign} \ T(g) \tag{OA7}$$

We examine sign $T(g)$ with respect to g. Note that from (OA2), for $w_1^* > 0$, we must have $\beta = \frac{\rho^2}{\delta}$ is sufficiently large or λ is sufficiently small.

Notice that:

$$T(0) = \lambda + c > 0 \tag{OA8}$$

$$T(0) = \lambda + c > 0 \tag{OA9}$$

Consider

$$T'(g) = \frac{\lambda\Delta}{(1+g)^2} - 2\lambda\Delta g + \frac{2\Delta g + \lambda - \lambda g}{(1+g)^3} +$$
$$c\left(\beta + \frac{\Delta}{(1+g)^2}\right)\left(\beta g - \frac{\Delta}{1+g} + \Delta + 1\right) \tag{OA10}$$

$$T''(g) = \frac{-2\lambda\Delta}{(1+g)^3} - 2\lambda\Delta - \frac{2\Delta(1-2g)}{(1+g)^4} - \frac{3\lambda(1-g)}{(1+g)^4} +$$
$$c\beta^2 + \frac{c\beta\Delta}{(1+g)^2} - \frac{3c\Delta^2}{(1+g)^4} - \frac{c\Delta(1+\Delta)}{(1+g)^3} \tag{OA11}$$

To show that $T'(g)$ has a unit root, we need to have $T''(g) < 0$ or $T'(g)$ decreases in g and has no more than one root for $g \geq \frac{\lambda}{\beta}$. For this to hold, we need to assume the following plausible assumption:

Assumption OA2. We either have a small probability of being detected (or a small c) or the NPO is highly risk-averse (or a large Δ). These assumptions are reasonable because the non-profit distribution assumption.

Using the L'Hopistal's Rule, we have $\lim_{g \to +\infty} T'(g) = -2\lambda\Delta \lim_{g \to +\infty} g + c(c + \lambda)\beta \lim_{g \to +\infty} g < 0$ (from Assumption OA2). As such, there must exist a unique root g_T of $T'(g)$.

Because $T'(g)$ decreases with g, $T'(g) > 0$ for $g \in (0, g_T]$ and $T'(g) < 0$ for $g \in (g_T, +\infty)$ or g_T is a local maximum of $T(g)$. Since $T(0) > 0$ we must have $T(g_T) > T(0) > 0$. Again, using the L'Hospital's rule and considering the limit of $T(g)$:

$$\lim_{g \to +\infty} T(g) = -\beta(\Delta - c) \lim_{g \to +\infty} g^2 < 0$$

That is, there exists a unique root of $T(g)$ such that $\tilde{g} \in (g_T, +\infty)$. To save space, we provide a graphical proof for ease of interpretation. Figure OA1 shows that, given $g \geq 0$ sign $T(g) < 0$ if and only if $g \geq \tilde{g}$ and sign $T(g) > 0$ iff $g < \tilde{g}$.

Following (OA7), sign $\frac{\partial b^*}{\partial g} < 0$ if and only if $g \geq \tilde{g}$ and sign $\frac{\partial b^*}{\partial g} > 0$ if and only if

$g < \tilde{g}$. Combining with (OA5), $\frac{\partial b^*}{\partial a^*}\Big|\partial g = \dfrac{\frac{\partial b^*}{\partial g}}{\frac{\partial a^*}{\partial g}} < 0$ if and only if $g \geq \tilde{g}$ and $\frac{\partial b^*}{\partial a^*}\Big|\partial g > 0$ if

and only if $g < \tilde{g}$.

The proof completes.

Figure OA1. How sign $\frac{\partial b^*}{\partial g}$ and sign T(g) behave when g varies in $(0, +\infty)$.
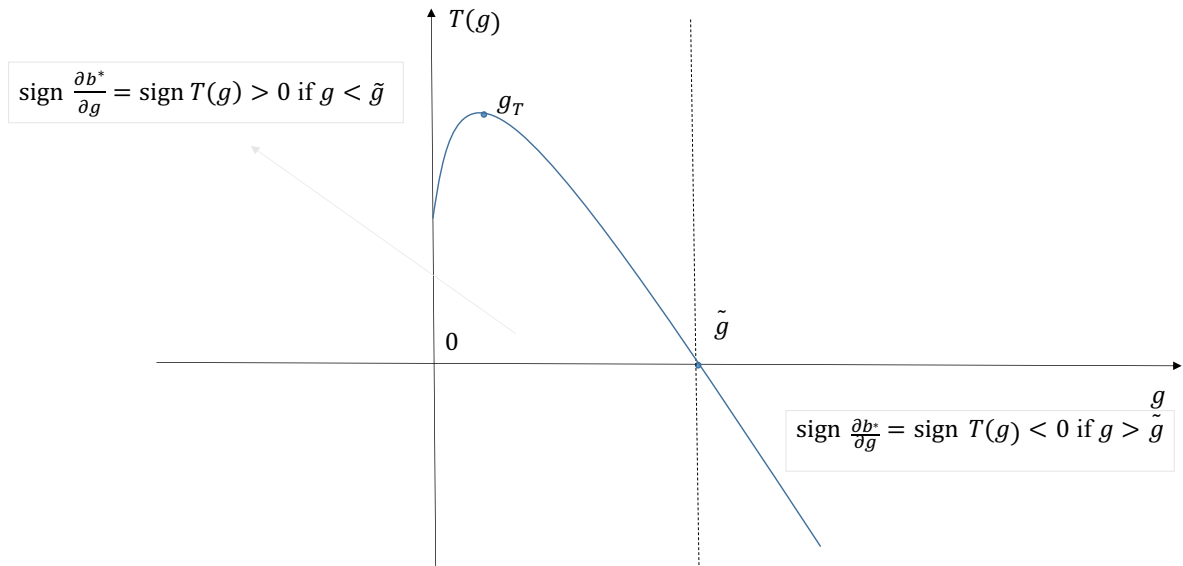
Note. The smooth line is for illustration only. The left-hand (right-hand) side $\tilde{g}$ represents the case when $T(g)$ is negative (positive, respectively). *Source*: Authors' own illustration.

2. Additional Tables

2.1. Robustness check: Replication for other measures of Governance Costs

As in Table A1, there are two categories underneath the broad item "Total Governance Costs": (i) Costs of Accounting and Audit Fees, and (ii) Costs of Administrative. In Table OA1, we replicate the main analysis replacing the aggregated variable *Governance Spending* by the two separate variables. The results remain qualitatively the same.

Table OA1. Replicated Results for The Other Measures of Governance Cost (Dependent Variable: MAD statistic)

VARIABLES	Costs of Auditing and Accounting Activities			Governance-Related Administrative Costs		
	(1) OLS with controls	(2) 2SLS	(3) Lewbel's (2012)	(4) OLS with controls	(5) 2SLS	(6) Lewbel's (2012)
Audit/Accounting Spending	2.658***	-1.146	3.055***			
	(1.000)	(3.589)	(0.806)			
Interaction Term 1	-5.770***	2.907	-6.082***			
	(1.650)	(7.345)	(1.284)			
Charitable Spending	3.047***	8.156***	4.674***	3.027***	9.221***	2.379***
	(0.566)	(1.936)	(0.735)	(0.565)	(1.930)	(0.688)
Interaction Term 2				-12.891**	-43.158	-14.842***
				(5.952)	(29.901)	(5.593)
Administrative Spending				10.601***	28.521**	10.782***
				(3.344)	(13.644)	(3.303)
Size (logged total assets)	0.472***	0.620***	0.593***	0.509***	0.672***	0.522***
	(0.081)	(0.094)	(0.079)	(0.083)	(0.099)	(0.080)
Age	0.081***	0.073***	0.077***	0.081***	0.071***	0.081***
	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)	(0.009)
Volunteers (number of volunteers)	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Being audited (1 = Yes, 0 = No)	-2.382***	-2.673***	-2.280***	-2.474***	-2.520***	-2.476***
	(0.468)	(0.496)	(0.462)	(0.464)	(0.468)	(0.461)
Receive government grants (1 = Yes, 0 = No)	-1.965***	-1.949***	-1.962***	-1.982***	-1.966***	-1.983***
	(0.264)	(0.266)	(0.263)	(0.263)	(0.265)	(0.262)
Zero fundraising (1 = Yes, 0 = No)	-0.212	0.272	0.035	-0.190	0.233	-0.172
	(0.271)	(0.306)	(0.268)	(0.271)	(0.308)	(0.270)
Losses from investments (1 = Yes, 0 = No)	0.356	0.369	0.274	0.349	0.369	0.365
	(0.346)	(0.351)	(0.342)	(0.347)	(0.350)	(0.345)
Receive restricted income (1 = Yes, 0 = No)	-0.576**	-0.565*	-0.557*	-0.534*	-0.519*	-0.507*
	(0.292)	(0.295)	(0.291)	(0.292)	(0.297)	(0.291)
Have endowment fund (1 = Yes, 0 = No)	-1.102***	-0.914***	-1.100***	-1.103***	-0.914***	-1.195***
	(0.345)	(0.354)	(0.343)	(0.345)	(0.355)	(0.345)
Income from Donations/Grants	-1.681***	-2.890***	-1.721***	-1.515***	-3.035***	-1.241**
	(0.582)	(0.816)	(0.593)	(0.576)	(0.768)	(0.573)
Number of non-zeros	-0.170***	-0.179***	-0.174***	-0.170***	-0.179***	-0.170***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)
Number of yearly reports	4.047***	4.340***	4.132***	4.054***	4.370***	4.033***
	0.472***	0.620***	0.593***	0.509***	0.672***	0.522***
Constant	43.531***	38.626***	40.725***	42.653***	36.932***	42.951***
	(1.471)	(2.130)	(1.464)	(1.509)	(2.311)	(1.524)
Observations	10,322	10,322	10,322	10,322	10,322	10,322
R-squared	0.262	0.243	0.261	0.263	0.249	0.262

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. *Interaction Term 1* is the interaction between *Charitable Spending* and *Auditing/Accounting Spending*. *Interaction Term 2* is the interaction between *Charitable Spending* and *Administrative Spending*. The unit of observation is individual British NPO

2.2.   Robustness check: Heterogeneity Analysis

Since our model predicts how an NPO behaves could vary by its reliance on different sources of income (such as government grants, restricted income, or endowments and donations), the effects of interest could be heterogenous over NPOs. To address this concern, we conduct two pieces of heterogeneity analysis. First, we divide our sample into subgroups by three binary variables, namely, whether: (i) the NPO receives government grants, (ii) the NPO receives restricted income, and (iii) the NPO has an endowment. We report the estimates for each variable in Panel A, B, C in Table OA2 respectively.

Second, we divide our sample into four subgroups of four quartiles of the *Income by Donation/Grants* variable and report the estimates for each quartile in Table OA3. Given government grants, restricted income and endowments only account for 38.4%, 48.1% and 16% of NPOs income we refrain from dividing the sample in this way for the actual (continuous) variables (see Table 2). Many NPOs receive zero income from these sources and dividing the sample by the actual incomes from these sources would lead to little variation in the subgroups. As reported in Table OA2 and Table OA3, even though the magnitudes are different (mainly due to different subsamples used), the signs and significance of the estimates of interest are largely in line with our main results. We use quartiles to split the sample into four groups here to retain substantial sample sizes for each subgroup, but (unreported) exercises using quintiles and percentiles also give qualitatively similar results.

Table OA2. Heterogeneity Analysis (Dependent variable: MAD_statistics)

Panel A. By whether the NPO receives government grants						
	Receive Government Grants = Yes			Receive Government Grants = No		
VARIABLES	(1) OLS	(2) 2SLS	(3) Lewbel	(4) OLS	(5) 2SLS	(6) Lewbel
Charitable Spending	4.374***	5.356	4.459***	3.334***	28.708***	1.649**
	(0.903)	(14.259)	(1.307)	(0.721)	(10.277)	(0.800)
Interaction Term	-20.020**	-139.365**	-22.597***	-17.046***	-51.112	-8.116
	(8.385)	(67.580)	(6.852)	(6.225)	(46.221)	(6.186)
Governance Spending	20.711***	76.396*	22.090***	7.485**	38.905*	2.129
	(4.868)	(40.287)	(4.421)	(3.504)	(23.492)	(3.564)
Observations	3,965	3,965	3,965	6,357	6,357	6,357
R-squared	0.242	0.158	0.242	0.237	-0.043	0.236

Panel B. By whether the NPO receives restricted income						
	Receive Restricted Income = Yes			Receive Restricted Income = No		
VARIABLES	(1) OLS	(2) 2SLS	(3) Lewbel	(4) OLS	(5) 2SLS	(6) Lewbel
Charitable Spending	6.631***	10.113	5.798***	2.124***	4.387	1.571*
	(0.782)	(12.069)	(1.207)	(0.742)	(6.954)	(0.833)
Interaction Term	-36.415***	-47.563	-35.863***	-9.243	-40.278	-11.698**
	(8.443)	(53.530)	(9.226)	(7.041)	(30.465)	(5.396)
Governance Spending	30.351***	38.475	29.784***	6.458*	20.951	7.176**
	(5.528)	(32.974)	(6.049)	(3.799)	(16.072)	(3.485)
Observations	4,965	4,965	4,965	5,357	5,357	5,357
R-squared	0.186	0.182	0.186	0.208	0.203	0.207

Panel C. By whether the NPO has endowment						
	Has endowment = Yes			Has endowment = No		
VARIABLES	(1) OLS	(2) 2SLS	(3) Lewbel	(4) OLS	(5) 2SLS	(6) Lewbel
Charitable Spending	4.787***	3.903*	1.934	3.059***	7.176	2.426***
	(1.204)	(2.344)	(1.708)	(0.664)	(6.057)	(0.758)
Interaction Term	-10.535	-13.291*	-0.832	-16.050***	-85.751*	-15.601***
	(11.965)	(7.097)	(9.047)	(5.962)	(51.835)	(5.390)
Governance Spending	7.434	7.923*	-0.287	11.836***	44.899**	11.256***
	(6.233)	(4.457)	(4.729)	(3.375)	(22.277)	(3.262)
Observations	1,653	1,653	1,653	8,669	8,669	8,669
R-squared	0.239	-0.273	0.235	0.262	0.244	0.262

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Robust standard errors in parentheses. All specifications include full controls as in Table 3. We replicate the results from the preferred specifications in the main analysis. The signs and significance of the estimates are generally similar to our main results.

Table OA3. Heterogeneity Analysis by quartiles of Income from Donations/Grants

		Panel A. Estimates for 1st and 2nd quartiles				

	1st Quartile			2nd Quartile		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	2SLS	Lewbel	OLS	2SLS	Lewbel
Charitable Spending	0.281	3.364	0.561	6.338***	-1.211	6.378***
	(0.788)	(3.956)	(0.836)	(1.002)	(1.331)	(1.760)
Interaction Term	2.618	-8.943	2.843	-31.142***	-7.158	-34.476***
	(7.620)	(147.037)	(5.861)	(10.109)	(5.068)	(8.955)
Governance Spending	-5.773	4.568	-6.716**	21.718***	2.250	23.186***
	(3.995)	(7.726)	(3.325)	(5.558)	(2.766)	(5.404)
Observations	1,653	1,653	1,653	8,669	8,669	8,669
R-squared	0.239	-0.273	0.235	0.262	0.244	0.262

		Panel B. Estimates for 3rd and 4th quartiles				

	3rd Quartile			4th Quartile		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	2SLS	Lewbel	OLS	2SLS	Lewbel
Charitable Spending	6.470***	1.703	4.331**	1.780	2.417	3.248
	(1.210)	(1.174)	(2.078)	(1.691)	(2.102)	(2.357)
Interaction Term	-29.695**	-13.788**	-20.115	-5.330	-22.613	1.254
	(12.222)	(5.357)	(12.302)	(12.376)	(28.018)	(10.194)
Governance Spending	24.346***	8.556**	15.423**	11.151	12.279	10.984*
	(6.945)	(3.581)	(6.884)	(7.014)	(11.870)	(5.687)
Observations	2,852	2,852	2,852	1,967	1,967	1,967
R-squared	0.213	0.176	0.212	0.274	0.121	0.273

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Robust standard errors in parentheses. The four quartiles are of Income from Donations/Grants. All specifications include full controls (except for the respective variable that the heterogeneity analysis is conducted on) as in Table 3. We replicate the results from the preferred specifications in the main analysis. The signs and significance of the estimates are generally similar to our main results.

Reference

Aebischer, P. (2012, January 19). Philanthropy: The price of charity. *Nature*, Vol. 481, p. 260. https://doi.org/10.1038/481260a

https://doi.org/10.1111/1475-679X.12058

Exley, C. L. (2019). Using Charity Performance Metrics as an Excuse Not to Give. *Management Science*, mnsc.2018.3268. https://doi.org/10.1287/mnsc.2018.3268

Garven, S. A., Hofmann, M. A., & McSwain, D. N. (2016). Playing the Numbers Game. *Nonprofit Management and Leadership*, *26*(4), 401–416. https://doi.org/10.1002/nml.21201

Gneezy, U., Keenan, E. A., & Gneezy, A. (2014). Avoiding overhead aversion in charity. *Science*, *346*(6209), 632–635. https://doi.org/10.1126/science.1253932

Hermalin, B., & Weisbach, M. (2012). Information Disclosure and Corporate Governance. *The Journal of Finance*, *67*(1), 195–233. https://doi.org/10.1111/j.1540-6261.2011.01710.x

SOPR. (2005). Accounting And Reporting By Charities: Statement Of Recommended Practice. Retrieved January 8, 2020, from 2005 website: https://www.gov.uk/government/publications/charities-sorp-2005