CREDIT Research Paper

No. 16/05

How Accurate are Financial Reports of British Charities?

by

Canh Thien Dang and Trudy Owens

Abstract

The demand for transparency and reliability of information from nonprofit organisations has surged as stakeholders increasingly use nonprofit financial information for contracting and regulating decisions. Contrary to previous studies, we theoretically show that attempts to enhance the monitoring of nonprofits through raising governance spending may not always lead to higher quality financial reports. Non-profits with larger charitable spending will be more inclined to report accurately, only if their spending on governance activities exceeds a certain threshold. To test the prediction, we measure the reporting quality at organisational level using public financial data from the UK Third Sector from 2007 to 2015. As popular measures require large and detailed data that are not usually available in the non-profit sector, we use Benford's Law to construct innovative and easy-to-replicate measures of information irregularities. Using these measures, we find robust results to support the theoretical predictions and address potential endogeneity by the conventional IV approach and Lewbel's (2012) heteroscedasticity-based IV estimator. This paper contributes to a heated debate in the UK regarding policies to govern the non-profit sector. We suggest that tighter monitoring might be ineffective to increase the sectoral transparency and accountability.

JEL Classification: L31 L44 D82

Keywords: British charities, non-profits, report accuracy, Benford's Law

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

CREDIT Research Paper

No. 16/05

How Accurate are Financial Reports of British Charities?

by

Canh Thien Dang and Trudy Owens

Outline

The Authors

School of Economics, The University of Nottingham; Authors can be contacted at Canh.Dang@nottingham.ac.uk and Trudy.Owens@nottingham.ac.uk

Acknowledgements

We thank Oliver Morrissey for his insightful comments and guidance, Raffaele Miniaci, Alex Possajennikov, Daniel Siedmann, Silvia Sonderegger, Richard Steinberg, Frank Windmeijer, participations at the ISTR Stockholm Conference and at the Nottingham PhD conference for their helpful feedbacks. Special thanks to Ronelle Burger for her ideas which helped shape the paper. Canh thanks the ESRC financial support for PhD study [grant number ES/J500100/1]; and the YITP of the IAAE at the University of Brescia. All remaining errors, as usual, are ours. The UK Third Sector Data collection is deposited at the UK Data Services.

Research Papers at www.nottingham.ac.uk/economics/credit/

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

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

I. Introduction

The growth and reputation of the non-profit sector throughout Western society has been tempered with scandals of fund misappropriation, abuse of power and lack of transparency. A large body of evidence indicates that the irregularities in the non-profit sector have become as relevant as they are in the corporate world.[1] Since a healthy non-profit sector is increasingly central to the well-being of society (Anheier, 2009) and an erosion of public trust can lead to reduced individual donations and public support for the sector (Steinberg, 2003), efficient and cost-effective way to asses the errors in organisation-level financial statement data of non-profit organisations becomes an important task for both regulatory bodies and academics studies.

For many organisations, keeping accurate and up to date financial records, however, is costly and time-consuming as it requires manpower and tools that otherwise could be spent on charitable missions. Quoting the Director of Acevo, Singh (2015), the UK's most influential network for Charity and Social Enterprise, captures the essence of the sector response: whilst transparency is necessarily a desired aim, spending on charitable services (usually known as programme ratio) needs to remain the priority. In addition to the expectations of givers that charities should consistently allocate most of their budgets towards charitable missions, there are perhaps two reasons for the sector's obsession with the programme ratio First, reporting a high programme ratio has become the most popular and appealing indicator of performance and financial efficiency that attracts future grants and donations (Garven et al., 2016). Second, since there is a lack of supporting academic evidence for the effectiveness of higher spending on governance activities on the sector's transparency, more resources spent in back offices could be unjustified and excessively expensive.[2]

We address two questions motivated by the issues: first, whether charities with higher key performance metrics are more diligent in their reporting quality; and second, whether charities with higher spending on back offices report more accurately?

Our contributions are threefold. First, we propose an innovative digital analysis based on Benford's Law to evaluate the extent of irregularities in financial reports using the UK Third Sector data from 2007 - 2015. Prior accounting literature outlines the drawbacks of current measures of reporting irregularities, including correlations with underlying firm characteristics and critical requirements of detailed time-

[1] Yetman and Yetman (2013) and Chen (2016) document that given financial reports are the leading indicator for donors and other stakeholders to evaluate organisational efficiency, some non-profits have misreported their finances to boost opportunities for donations and volunteering supply. The Centre for Policy Studies, a British think tank, reproduces the accounts of 50 of the largest UK charities by income and suggests that the actual income figures of these charities could be more than double their self-reported accounts (Norton, 2014). Keating et al. (2008) find that 74% of the regulatory filings from American non-profit organisations fail to properly report categorical expenses, though the authors consider the inaccuracy can partially be due to unintentional errors. Beyer et al. (2010) provide a comprehensive review in the corporate literature; for non-profit studies, see Hofmann and McSwain (2013).

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

series, long panel or forward looking data (see Dechow et al., 2010). Such drawbacks hinder the use of those measures in the non-profit context as the literature has long been cited as lacking of data. Following studies in the statistics and mathematics literature regarding Benford's Law (see Amiram et al., 2015), we construct easy-to-replicate measures of financial statement inaccuracy by assessing the distribution of the first digits (e.g., the first digit of the number 348.79 is 3) of all the numbers in a non-profit's annual financial datasets. The approach has several key advantages. First, the existing financial statement data in public domains are sufficient to implement the assessment, with no additional data collection needed. Second, by focusing on the statistical distribution of the first digit of the numbers in a dataset, our approach separates confounding factors such as forward-looking information or managers' characteristics. Third, the approach is flexible in the sense that it can be automated and can be easily extended to a more general setting with other augmented data. We should emphasise that although the calculation indicates irregularities in some organisations' data, the approach is not fail-proof, nor will it substitute existing methods based on auditing. Nonetheless, we believe it can serve as a useful and relatively low-cost first step for effective and more targeted auditing.

The approach is motivated by a similar idea in a statistical analysis by Jacob and Levitt (2003), who use unexpected test score fluctuations and suspicious patterns of student answers to detect teacher cheating. Some organisations may undertake sophisticated activities that allow errors in their financial statements to remain undetected by regulating bodies (the Charity Commission in the UK or the Internal Revenue Service in the US); such activities, however, are also likely to leave telling signs of errors in the forms of invalidating the distributional properties of true numbers of financial data. Informally, Benford's Law states that the first non-zero digits of all the numbers contained in certain empirical datasets are not uniformly distributed – as one may naively expect – but follow a logarithmic-type decreasing frequency known as Benford distribution (that is, 1 will appear as the first digit 30.1% of the time, 2 will appear 17.6% of the time, and so forth).[3] Any unexpected deviations from the distribution are flags of potential errors.

Hill's (1995) theorem formalises the idea: if distributions are randomly selected and samples are taken from each of these distributions at random, the combinations of these samples tends toward the distribution predicted by Benford's law. These conditions are likely to apply to accurately reported financial data. The true (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 and after a given period. These interactions could be considered as randomly distributed because they are unknown to other individuals except those who are involved. The financial items representing these interactions, therefore, are estimates of the realisations from unknown and random distributions. Since different financial items are likely to be determined by different mechanisms

[3] A related distribution that is 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).

(for example, the distribution of revenues from government funding is likely to differ from that of administrative costs), the mixture of these distributions, which constitute an organisation's financial report, may follow Hill Theorem. The distributions of the first digits of all numbers in the financial reports is then expected to follow the Benford distribution. Specifically, the aggregated set of numeric items representing revenue sources from grants, businesses or investments, together with expenditure on salaries, charitable activities, taxation, etc., is expected to follow Benford's law. To construct our measures of irregularities, we exploit a plausible assumption that since these financial items are self-prepared and historically less attentive in the non-profit sector, some preparers might introduce errors that make the dataset deviate from Benford's law, whether in the form of human mistakes, biases or manipulation.[4]

We proximate the degree of inaccuracy of each non-profit (hereafter NPO) by the extent that the distribution of the leading digits of figures in the organisation's self-reported annual financial statements diverges from the theoretical Benford distribution.[5] While earlier studies use Benford's law to provide descriptive evidence of irregularities in financial data (see Amiram et al., 2015), macroeconomic data (Michalski and Stoltz, 2013) and Belgian non-profit data (Van Caneghem, 2015); our paper represents the first systematic attempt to (1) identify the overall prevalence of financial report errors, and (2) analyse the factors that predict the errors. At the organisational level, we find that 25% of our sample charities provide financial figures that significantly deviate from the Benford distribution, indicating possibility of errors.

Second, we develop a simple model to parametrise how an agent (NPO) chooses the level of misreporting when reporting to a principal (donor) in a three-period agency set-up. The donor contracts the NPO to undertake a development project whose ultimate value is only realised in the terminal period. In the initial period, the NPO chooses their optimal level of effort and the degree of inaccuracy (inflated bias) to induce in a report about the project's intermediate value to the principal. We implicitly assume that the NPO can influence the overall report accuracy through either increased diligence or integrity; however, the agent faces a cost of misreporting information due to some exogenous governance constraints (such as the accounting environment or auditing requirements from the donor). In the second period, the NPO privately observes the intermediate state of the project, which is influenced by their chosen effort and medium-run stochastic events, and reports the state to the donor. Having a commonly known prior belief of the agent's information in equilibrium, the donor forms their belief by deducting the prior belief from the (potentially misleading) report. The donor then pays the contracted grant which is linearly proportionate to this belief of the project intermediate value. In the long-run, terminal period,

[4] Hal Varian (1972) in promoting the use of Benford's Law in economics suggests that the preparers may be biased towards simpler or more intuitive distributions such as the uniform distribution.

[5] As such, the degree of inaccuracy is defined as the proportion of a given report being verified as inaccurate. Another popular term in the accounting literature is credibility of information disclosed (see Healy and Palepu, 2001 for a review). In other words how confident you are that the report is accurate. We use a Chi-square test to compare the observed distribution of the first digits with the theoretical distribution to capture this interpretation of informational accuracy.

all stochastic events are realised and the true project's value is recognised. The donor now earns their payoffs as the terminal value net of the grant payment. The donor aims to maximise the net terminal social impact in the long run; while the NPO aims to maximise its utility derived from the grant received and the efforts exerted to influence the project's value and the report accuracy.

Our predictions are as follows. The NPO will always report accurately either when the first-best solution is possible (the project' terminal value is contractible) or when their incremental reputation concern regarding the donor' prior belief of the agent's misreporting in equilibrium is relatively high. Otherwise, there exists some inaccuracy in optimum. The correlation between the optimal amount of misreporting and the optimal amount of exerted action depends on the cost of information misreporting compared to an exogenously set threshold. When the governing environment is relatively relaxed, NPOs exerting greater effort also tend to exaggerate more. When the environment becomes stricter, the NPOs would expend resources to deliver higher effort rather than misreporting. The effect of the cost of information misreporting on the amount of irregularities also depends on the relative performance of the NGO. Only if the agent optimally chooses effort that exceeds a threshold (for example, a sectoral norm), higher cost of information misreporting will be associated with a lower level of reporting inaccuracy.

Finally, we test the theoretical predictions by examining the effect of the reported fraction of income spent on charitable activities, a proxy for the optimal effort, and the fraction of income spent on governance activities on the reporting accuracy as a proxy for the cost of misreporting information. We propose an IV identification strategy to control for potential endogeneity of the observed charitable spending ratio. Previous studies show that this ratio is susceptible to misstatement or strategic manipulation by the report preparer. This endogeneity problem has remained largely unaddressed in non-profit studies. We first rely on the exogeneity of the number of staff and the recorded spending on social security, which are either easily verified or publicly recorded. These instruments are strongly correlated with the charitable spending ratio and we expect these variables remain orthogonal to the level of reporting inaccuracy. Even if the orthogonality fails, we compliment the traditional IV approach by the heteroscedasticity-based estimator proposed by Lewbel (2012), which does not rely on the standard exclusion restrictions.

We empirically find robust and supportive results for our theory. Increased charitable spending (or programme ratio) leads to more accurate financial reports if the NPO spends at least 15%-40% of their income on governance activities. If the threshold is not met, the NPO may either exaggerate or neglect their reporting activities. On the other hand, increased accuracy of financial reports is associated with increased spending on governance activities only when the NPO already spends at least 70%-75% of their total income on charitable activities. If the threshold is not met, larger governance spending may put pressure on the organisation to misreport their data. Moreover, inaccuracy appears to be systematically higher in cases where the costs of preparing accurate financial numbers are higher (e.g., in larger and older charities in which the loads of accounting tasks become more complex over times), or the probability of errors being detected is higher (the financial reports being audited or receiving government grants). Having restricted income or endowment funds are also positively correlated with more accurate financial information.

For sensitivity analysis, we first show that our results are not sensitive to the constructing algorithms of our measures. We alternatively use four indices that measure deviations from a reference distribution: (i) MAD (Median Average Deviations), (ii) (the Chi-square test statistics of goodness of fit), (iii) the Kolmogorov – Smirnov (KS) statistics KS statistics, and (iv) a binary variable *Deviate* indicating whether the non-profit's data deviate from the Benford distribution using KS test at 10% of significance. To address the possibility that the errors detected by our digital analysis are not driven strategic behaviours but simply by poor book-keeping, we replicate our results even after excluding NPOs with spending on governance activities (accounting and administrative activities) in the bottom 10%, 25% or 50%. We also replicate the results after controlling for the impact of non-disclosure when some NPOs may record zero transactions so that there is no information for our digital analyses by using Heckman sample correction model.

We present the theoretical model in Section II. Section III explains Benford's Law, the UK Third Sector data and how our proxies are constructed. Section IV describes the traditional IV approach and Lewbel's (2012) heteroscedasticity-based IV estimator. Section V presents our findings. We check the sensitivity in Section VI. Section VII concludes.

II. A model of the optimal information (in)accuracy

Consider a three-period reporting game between a donor (principal) and an NPO (agent). The set-up is an adaption of Goldman and Slezak (2006) and similar to Crocker and Slemrod (2007) in the sense that the NPO may take a hidden action that affects the (actual) terminal social impact of the project. When partly realised in the intermediate period, the intermediate value constitutes hidden information that is privately observed by the NPO who may then issue an inflated report. We aim to provide a model that specifies an optimal contract between the principal and agent that not only provides the agent incentive to enhance the project's actual ultimate value, but also minimises the agent's incentive to misreport. Note that it remains possible that deviating from the true value can be caused by factors other than intentional manipulation, such as failing to comprehend/estimate the current state of the organisation or simply human error in information recording. Such cases are difficult to distinguish in the current theory. Throughout the analysis, we implicitly assume these other factors are captured in an organisation-specific stochastic error term. We henceforth refer to deviation from the true state of the project as information inaccuracy (or misreporting) instead of information manipulation to emphasise that the notion also includes other types of deviation from the truth without intention to deceive.

1. Action, intermediate report, misreporting and terminal value

At $t = 0$, a risk-neutral donor contracts with a risk-averse NPO to deliver a social project that yields a terminal value in the long run $t = 2$.[6] During the initial period, given the grant conditions, the NPO privately makes a one-time decision of (1) the amount of unobservable action $a \geq 0$ (such as the level of dedication or effort to exert), and (2) 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-run state of the project). Exerting productive effort and producing biased reports 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 $\psi^b(b) = \frac{g}{2}(b - \tau)^2$, where $g > 0$, cover the NPO's cost of producing a report (g) with an amount b of misreporting and a reservation level of misreporting (τ). The misreporting cost may reflect the time spent lobbying the auditor or coming up with creative ways to make the report more attractive to the donor. When g is common knowledge, we interpret g as an observable organisation characteristic representing the NPO's *governance structure*. The *governance structure* is used to reflect a broad sense of the donor-NPO information environment such as regulatory technology (i.e., the expected value of any penalties imposed on the NPO), board/committee composition or the NPO's accounting divisions and auditors.[7] We interpret *governance* as a mechanism, possibly required by the donor, for aligning the interests of the NPO more closely with those of the donor and *governance structure* (*g)* is the cost to achieving this alignment (see Thakor, 2015 for a similar argument). The parameter τ refers to a reservation amount of misreporting at which the NPO would incur no cost. We assume $\tau \geq 0$ to capture the idea that providing an accurate report ($b = 0$) may be costly to the NPO. For example, the NPO must spend resources of $\frac{g}{2}\tau^2$ on hiring highly trained accountants to have well prepared financial reports. Finally, the NPO incurs a reputation loss $\psi^c(b)$ for deviating from the donor's prior belief of the NPO's equilibrium extent of misreporting (discussed below).

The chosen action a at $t = 0$ induces a gross terminal value of the project realised at $t = 2$, denoted as V, according to $V = \rho a + \eta + \varepsilon$. 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. Both η and ε are realised only at $t = 2$ and remain unknown to the NPO when choosing action a at $t = 0$. We assume ρ, σ_η^2 and σ_ε^2 and the parameters of the cost functions are common knowledge.

At $t = 1$, the NPO privately observes the intermediate state of the project and issues a (potentially misleading) report θ about the state to the donor. The intermediate state can be thought of as the outcome of action a coupled with fundamental economic events that happen to the NPO during $t = 1$. Denote θ^T the true signal of the state. We assume that θ^T perfectly captures $\theta^T = \rho a + \eta$, that is, given the donor

[6] Wedig (1994) argues that nonprofits typically are risk-averse due to their non-distribution incentive.

[7] We follow Beyer et al. (2014)'s interpretation. When g is private information, it can be interpreted as the agent's integrity (intrinsic aversion to lying, see Gneezy, 2005) and the innate capability to manipulate. Beyer et al. (2010) review current research and suggest that the information environment (accounting structure and stewardship) is as relevant and applicable in the non-profit sector as in the for-profit sector.

receives the true signal θ^T, the donor can form a rational belief about the gross terminal value of the project, i.e., $E[V|\theta^T] = \theta^T = \rho a + \eta$. The true signal θ^T differs from the realised gross terminal value only by the random idiosyncratic shock ε, which reflects all the remaining uncertainty that cannot be observed given the intermediate state.

If the NPO truthfully reports their state, $\theta = \theta^T$. Otherwise, by construction, the NPO misreports by an amount b and issues $\theta = \theta^T + b$. Assume for convenience that $b \geq 0$, so the NPO always tends to over-report the project's gross terminal value.[8] Based on the observed report, θ, the donor forms their expected terminal value of the project, S, and disburses the contracted grant W.

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 for activities to prepare for the misstated report at $t = 1$ (e.g., monetary cost for bribing/colluding with auditors or the project's opportunity cost of the NPO's time spent preparing for manipulative activities). This diverted resource reduces the project's terminal value. For simplicity, we assume that the amount of the bias is linearly correlated with the diverted resources. That is, for an amount of bias 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:

$$V_2 = V - \lambda b = \rho a + \eta + \varepsilon - \lambda b = \theta^T - \lambda b + \varepsilon \quad (1)$$

2. Payoffs and the optimal contract

The first-best solution occurs when the donor could solely contract on the project's terminal value (by construction, not subject to misreporting). The agent's action is contractible and there is no misreporting. In most cases this solution is not realistic because social (development) projects' ultimate impact takes time to realise, often so far in the future that the NPO's compensation cannot wait until the gross terminal value is recognised.

Instead, we focus on the second-best solution in which there exists hidden action and information. The donor needs some indication about the project's terminal value in order to measure the NPO's performance and to pay out the contract at $t = 1$. The expected social value S, based on the self-report θ, becomes the only observable performance measure the donor can rely on for contracting with the NPO. Thus, we assume that the grant $W(S)$ depends on the donor's intermediate period's belief about the project's terminal value. For tractability, we assume that at $t = 0$ the donor designs a funding contract $W(S)$ that is linearly correlated to the expectation, S, as:

$$W(S) = w_o + w_1 S \quad (2)$$

where w_o represents the project's endowment and w_1 reflects the value-sensitivity of the contract.

[8] This assumption is consistent with manipulation incidents discovered in the sector. Krishnan and Yetman (2011) find that non-profit hospitals in California report upward-manipulated program ratios to the state regulatory agency by 8%.

As the donor is aware of potential misreporting and the diverted resource cost, given the report θ, the donor forms their belief based on the net terminal value V_2. Formally, at $t = 1$, the expectation of the net terminal value given the report θ is:

$$S = E[V_2|\theta] = E[\theta^T - \lambda b + \varepsilon|\theta] = \theta - b^e - \lambda b^e = \rho a + \eta + b - b^e - \lambda b^e \qquad (3)$$

where b^e is the donor's prior belief about the equilibrium amount of misreporting. Following Stein (1989) and Goldman and Slezak (2006), we assume that this prior belief b^e is formed before the contract emerges and remains exogenously fixed due to the one-off nature of the interaction.[9] This belief could be formed prior to the contract by examining the NPO's records and their organisational structure and forming an expectation regarding the equilibrium level of misreporting that the NPO undertakes. We assume that this belief is fixed during the game timeframe and there is no updated posterior belief after observing the report θ. When this belief is rational (such that the records are verified so that the donor perfectly predicts $b = b^e$ in equilibrium and $b^e = b^*$ where b^* denotes the equilibrium level of misreporting resulting from the optimal contract), information misreporting has no impact on the rational expectation of the gross terminal value *in equilibrium.* Otherwise, the expectation S is increasing in the actual amount of misreporting b and decreasing in the expected extent of the NPO's misreporting b^e. To study the effect of the commonly known, fixed prior belief on the NPO's optimal manipulation, we introduce a reputation concern function into the NPO's payoffs $\psi^c(b) = c(b - b^e)$. The linear functionality is for tractability and captures two notions. First, the reputation loss is decreasing in the actual amount of misreporting, regardless of the prior belief b^e. Second, if the NPO's misreporting choice is *better* than expected, that is the NPO misreports less than what the donor expects before the game $b < b^e$, the NPO's reputation gains $\psi^c(b) > 0$. Otherwise, the NPO's reputation loses $\psi^c(b) < 0$. The parameter $c \geq 0$ represents the degree of how deviating from some prior belief of their reputation matters to the NPO.

At $t = 1$, the NPO undertakes the contract, receives $W(S)$ and incurs costs $\psi^a(a) = \frac{\delta}{2}a^2$, $\psi^b(b) = \frac{g}{2}(b - \tau)^2$ and reputation concern $\psi^c(b) = c(b - b^e)$, the induced wealth of the NPO is:

$$\omega = W(S) - \psi^a(a) - \psi^b(b) - \psi^c(b) \qquad (4)$$

We further assume that the NPO has a constant absolute risk aversion (CARA) utility function, $u(\omega) = -\exp(-r\omega)$, that is, a negative exponential von Neumann-Morgenstern utility function with the Pratt-Arrow absolute risk aversion coefficient of $r > 0$. Without loss of generality, we assume that the NPO's reservation utility u_0 is zero.

Lemma 1. Given the utility function $u(\omega)$ and the NPO's information set Ω_0^N at $t = 0$, the NPO would require a certainty equivalent for undertaking a contract W whose induced wealth at $t = 1$ is given by (3) as:

[9] One extension to this model is 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.

$CE(W, a, b|\Omega_0^N)$

$$= w_0 + w_1(\rho a + \eta + b - b^e - \lambda b^e) - \frac{\delta}{2}a^2 - \frac{g}{2}(b-\tau)^2 - c(b - b^e) - \frac{r}{2}w_1^2\sigma_\eta^2 \quad (5)$$

Proof. See Appendix A.

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

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 function by choosing action a and the misreporting amount 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 with respect to a and b:

$$\max_{a,b} CE = w_0 + w_1(\rho a + \eta + b - b^e - \lambda b^e) - \frac{\delta}{2}a^2 - \frac{g}{2}(b-\tau)^2 - c(b - b^e) - \frac{r}{2}w_1^2\sigma_\eta^2 \quad (6)$$

The first-order condition with respect to a and b gives the NPO's action and misreporting choice as in Corollary 1.

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

$$a^* = \frac{\rho}{\delta}w_1 \ (7)$$

$$b^* = \max\left\{0, \frac{w_1 - c}{g} + \tau\right\} \ (8)$$

As standard in the literature, to induce the agent to exert any action, the principal must offer a contract that is sensitive to the performance measurement S ($w_1 > 0$); while the optimal action is decreasing with the marginal rate of disutility δ. The positive sensitivity, however, can lead to positive misreporting if the incremental reputation concern c is sufficiently small relative to the NPO's incremental value-based sensitivity w_1. When the reputation concern, for example the NPO wants to maintain a good record or future contract with the donor, is sufficiently high so that it dominates the marginal benefit from the performance-based sensitivity, the agent's optimal strategy is to report truthfully. For the remainder of the analysis, we examine the situation that leads to positive manipulation, where $w_1 > c - \tau g$.

The donor considers the optimal amount of action and misreporting to design a contract $W(S)$ that maximises the net terminal value minus grants paid to the agent. Formally, the donor solves the following problem:

$$\max_{w_0, w_1}\{V_2 - W(S)\} \quad (9)$$

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

$$CE = w_0 + w_1(\rho a + \eta + b - b^e - \lambda b^e) - \frac{\delta}{2}a^2 - \frac{g}{2}(b-\tau)^2 - c(b - b^e) - \frac{r}{2}w_1^2\sigma_\eta^2 \geq u_0 = 0 \quad (10)$$

At equilibrium, the participation constraint holds at equality, that is, $CE = 0$, implying that:

$$w_0 = -\left[w_1(\rho a^* + \eta + b^* - b^e - \lambda b^e) - \frac{\delta}{2}a^{*2} - \frac{g}{2}(b^* - \tau)^2 - c(b^* - b^e) - \frac{r}{2}w_1^2\sigma_\eta^2\right] \quad (12)$$

Substituting w_0, the optimal amount of action (8) and misreporting (9) into the maximisation problem

(10) yields:

$$\max_{w_1} \frac{\rho^2}{\delta} w_1 - \lambda[\frac{(w_1 - c)}{g} + \tau] - \frac{\rho^2 w_1^2}{2\delta} - \frac{(w_1 - c)^2}{2g} - c\left[\frac{w_1 - c}{g} + \tau - b^e\right] - \frac{r}{2}\sigma_\eta^2 w_1^2 \quad (13)$$

The first-order condition gives the unique equilibrium contract (w_0^*, w_1^*) in Corollary 2.

Corollary 2. There exists a unique (w_0^*, w_1^*) such that w_1^* satisfies (14) and w_0^* follows (12):

$$w_1^* = \frac{\frac{\rho^2}{\delta} - \frac{\lambda}{g}}{\frac{\rho^2}{\delta} + \frac{1}{g} + r\sigma_\eta^2} \quad (14)$$

In line with the literature, we assume that the exogenously set *governance structure* g is sufficiently high, $g > \frac{\delta\lambda}{\rho^2}$, so that the performance-based sensitivity w_1^* is positive. In this case, the equilibrium performance-based sensitivity is a function of the diverted resource cost λ, the governance structure g, the incremental compensation for productivity ρ, action cost δ and the organisation-specific risks borne by the NPO $r\sigma_\eta^2$.

Substituting w_1^* into (8), the optimal amount of misreporting now becomes:

$$b^* = \frac{\beta - \frac{\lambda}{g}}{g\beta + g\Delta + 1} - \frac{c}{g} + \tau \quad (15)$$

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

3. Comparative statics

The following propositions, implied by Corollary 1 and 2, specify comparative statics that characterise the equilibrium interaction between the optimal amount of action and misreporting with respect to observable characteristics. We restrict the results of interest for our empirical analysis below.

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

$$\tilde{g} = \arg\max_{\mathrm{g}} b^* = \frac{\lambda + c + \sqrt{(\lambda + c)\lambda + \frac{\beta(\lambda + c)}{\beta + \Delta}}}{\beta - c(\beta + \Delta)} \quad (16)$$

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

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

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

Proof: See Appendix B.

We can interpret Proposition 1 as follows. Part (i) suggests that if an NPO's current governance structure remains below a certain threshold, the NPO responds to their higher optimal choice of action by increasingly exaggerating the reported level of their impact. If the governance structure passes the certain threshold, Part (ii) suggests that a greater choice of action to exert is accompanied with lower degree of inaccuracy or higher report reliability. Part (i) suggests that if the organisational governance is not effectively strict, the NPO could inflate their reported impact following their higher level of action. These NPOs may believe misreporting behaviour is tolerable (or effective) when the governance structure is set below a threshold, otherwise they would abstain from such behaviour when their optimal action is high, as in Part (ii). Another interpretation is that the higher degree of inaccuracy following increased charitable effort may be the result of less diligence in reporting. When the governance spending is low so that both the donor and the NGO agree on the importance of monitoring activities and the priority of the project, the NPO will prioritise the project and neglect the reporting task. The increased charitable effort is then associated with a report more prone to errors. It is the trade-off between spending on actions that improve the project's ultimate value, and accountability of any sequential reports, which results in the increasing tolerance for minor digressions. Part (ii) suggests that when the governance environment places stricter importance or higher weight on reliable reports, the NPO would spend time and effort improving their optimal action choice rather than finding creative ways to manipulate their reported values. The stricter environment, in fact, may effectively prevent manipulative behaviour, inducing the NPO to focus more on impact-enhancing activities rather than information inflating. A pessimistic view is also possible. If governance becomes excessively strict, such as when the donor visits the project too often or requests too many reports/assessment meetings, the NPO responds to the contract with a lower optimal choice of effort and increases their manipulation level b^*.[10] The following proposition specifies how the governance structure impacts on the optimal extent of manipulation in accordance with the chosen optimal action.

Proposition 2: When the exogenously set governance structure is sufficiently high for the donor to offer a positive performance-based sensitivity $w_1^* > 0$, there exists a fixed threshold of the optimal action $\tilde{a} > 0$ such that:

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

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

Proof: See Appendix C.

Proposition 2 suggests the interdependence between the optimal choice of action and the exogenously set governance structure. If the NPO's optimal level of action is lower than an exogenously predetermined threshold, $\tilde{a}$, the stricter governance structure will lead to more manipulation. In contrast, if the optimal action is exceptionally high, the stricter governance structure becomes effective as the manipulation

[10] If we interpret spending on governance as an agreement between the NPO and the donor (lowered governance means higher agreement in our context), Proposition 1 is consistent with Thakor's (2015) model of strategic information disclosure when there is fundamental disagreement regarding interpretations of information disclosure between a firm and its investors. He shows when investors (donor) are in higher agreement with the organisation's intrinsic value, a more valuable firm (higher programme ratio in our context) discloses less information and vice versa. One intuition is that greater information disclosure may reveal the firm's strategy and this transparency makes the firm fragile so that both investors and the firm tend to agree with opaque disclosure.

extent of sequential reports is decreasing. The threshold could be held by the donor prior to the contract based on a common norm/public trust on how NPOs are expected to perform in a particular charitable activity. The intuition becomes clear if we interpret the threshold as classifying two types of agents: high (low) type NPOs are those who choose their optimal effort more (less) than an industry-imposed threshold. Low-type NPOs could expect greater pressure under a more demanding governance structure as they are underperforming according to the threshold expectation. Those that are well below the expected level may respond by exaggerating the project value more than those just below the expected level. On the other hand, the high-type NPOs, which are doing well by exceeding the industry expectation, may produce even more reliable reports following stricter governance requirements. One reason may be that as the higher NPO type is now more concerned with their reputation, they respond more vigilantly to even higher pressure from governing bodies.[11]

Using the first-order condition for Equation (15), we specify other comparative statics in Corollary 3.

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

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

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

Part (i) suggests that NPOs with higher productivity, higher resources lost due to manipulation and higher reputation concerns will choose a lower level of optimal misreporting. In contrast, the more it costs to produce a report free of errors, τ, the greater the extent of misreporting incidents, possibly white-lie errors. Part (ii) implies that the higher cost of exerting action is correlated with greater information manipulation; whereas higher risk aversion and higher organisation-specific risk variances induce larger information inaccuracy. The intuition is that an NPO may opt to manipulate activities instead of spending increasingly costly effort in improving the terminal value. In contrast, the NPO would respond to a more volatile environment by inflating their report, probably in the hope that the donor will mistake the inflated information as organisation uncertainty (η). The common underlying explanation for these statics is due to the contract structure that pays out the compensation before the verifiable terminal value is realised. The contracted agent, therefore, faces a trade-off between expending efforts to improve the true state of the project and manipulate the report on which the contract is based.

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

1. Forensic Economics Studies and Benford's Law

Measuring irregularities in economics data is a well-researched area (see Zitzewitz, 2012). The for-profit literature has attempted to construct and validate measures of reporting manipulation such as Benesish's

[11] Part (i) result is consistent with Thakor's (2015). He shows that it is possible for firms with improved corporate governance to disclose less information in equilibrium (hence larger bias b in our case).

M-score, accrual-based estimates from models (see Jones, 1991), earnings management (see Dechow et al., 2010) or distributional analyses (Burgstahler & Dichev, 1997). Rather than focusing on accruals and earnings, non-profit studies often attempt to measure irregularities in reports by estimating expected programme ratios (Trussel, 2003), levels of charity care (Vansant, 2011) or fundraising and administrative expenses (Yetman and Yetman, 2011). There are, however, weaknesses inherent in these measures. First, measures estimated from prediction models incur sample selection bias and measurement errors (Dechow et al., 2010). Second, these measures require strong assumptions about the organisations objective function and managers' incentives, which are not always realistic and could induce correlation between the measures and the organisation's characteristics (see Amiram et al., 2015). Third, these models require forward-looking information and often detailed time-series and panel data. This requirement tempers their usefulness in non-profit studies, in which small sample size and data irregularities are often the main challenges.[12]

We use an alternative proxy for measuring accuracy of a self-reported set of financial data based on Benford's Law. Benford's Law, also called the first-digit law, is a mathematical law regarding the frequency distribution of leading digits in many sets of numerical data (e.g., the leading digit of the number 1201.17 is 1). Contrary to basic intuition, the occurrence of each digit as a leading digit in a set of numbers is usually not equal. Newcomb in 1881 discovered that the first digits of all numbers in an empirical dataset will occur with a logarithmically decreasing frequency. Benford in 1938 published a series of datasets that adhere to the law. The theoretical foundation is based on Hill's (1995) theorem, which states that if distributions are non-truncated or uncensored, random samples of varying magnitudes taken from a random mixture of those distributions will have the first digit converging to the logarithmic of the Benford distribution. Hill's (1995) theorem provides the following formal derivation of the distribution according to Benford's Law:

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

where $P(d)$ is the probability that digit $d = 1, 2, \ldots, 9$ occurs as the leading digit in a naturally drawn set of numbers. Table 1 records the full theoretical distribution specified by Benford's Law: 1 will appear as the leading digit 30.1% of the time, 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 intuition behind why accurate empirical data 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 log and obtain the fraction behind

[12] Hofmann and McSwain (2013) reviews the challenges in non-profit studies. Amiram et al. (2015) elaborate weaknesses of these measures in the for-profit literature.

the integer. For example, if the fraction is between 0 and 0.301, the first digit of *N* is 1. If the fraction is between 0.301 and 0.477 (interval of 0.176), the first digit is 2 et cetera. It implies that the intervals between the fractions of the decimal point of the log number ($\log_{10}(1+\frac{1}{d})$) is equivalent to the probabilities that digits appear as the leading number. Second, if the probability distribution function of the logarithm of *N* is smooth and symmetric, a number will be in the internal between *n* and *n + 0.301,* where *n* is any integer in the logarithmic distribution, with a probability of 30.1%, between *n + 0.301* and *n + 0.477* with 17.6% *et cetera.* Third, due to the Central Limit Theorem, distributions drawn from a random mixture of different distributions tend to be smooth and symmetric and hence Benford's Law applies to sets of data that comprise of different sources of numbers.

Since the discovery of Benford's Law, numerous empirical studies have emerged both to verify the law in specific datasets and to detect errors and frauds in publicly available data (see Kossovsky, 2015 and Nigrini, 2012 for fuller reviews). Examples include suspicious national statistics (see Rauch et al., 2011 for an application on EU-governmental data), questionable election and survey data or manipulated regression results in published scientific studies (see Judge and Schechter, 2009 for another review). Literature in statistics and mathematics has also come to a consensus that comparing the observed distribution of the leading digits in a dataset with its theoretical distribution according to Benford's Law allows users to evaluate the level of reporting errors within the underlying data (see Miller, 2015 for complete proofs and applications).[13]

In economics, accounting and finance, the idea that Benford's Law could be used to detect errors and manipulation in economic data was first suggested by Varian (1972).[14] Hill's (1995) theorem supports this suggestion as economic/accounting data are a series of estimations of economic activities (e.g., cash flows from sales, payments or expenses) with distributions which are both generated from different mechanisms and can vary in magnitude. For example, the distribution of cash flows from administrative costs that occurred during a fiscal year is possibly different from the distribution of that from grants received. The mixture, usually random and unknown, results in distributions underlying the financial figures which may indeed follow the criteria in Hill's (1995) and Benford's Law. Durtschi et al. (2004), Michlski and Stolz (2013), and Amiram et al. (2015) discuss how Benford's Law can be effectively used by auditors in detecting errors and frauds in annual reports and macroeconomic data. Busta and Weinberg (1998) perform an analytical review on simulated datasets to conclude that digital analysis based on Benford's Law outperforms most other measures of accounting quality. Amiram et al. (2015) provides the first simulated analysis using stylised financial statements to ascertain the empirical distribution of

[13] In the non-profit literature, comparing empirical distributions of a pool of accounting information to a theoretical distribution is not unprecedented. Bhattachary and Tinkelman (2009) examine GuideStar data of 111,000 non-profits by distributional analysis and find no evidence of expense allocation manipulation. Ballantine et al. (2007) use the same method and find highly significant discontinuity in residual incomes of English NHS hospitals during 1998-2004.

[14] Following Zitzewitz's (2012) taxonomy, measures based on Benford's Law can be categorised as a statistical model-based approach. One example in the same spirit is Jacob and Levitt (2003) in which they derive a testing procedure for teachers' teaching. The main assumption is similar to ours: fraudulent cases exhibit patterns that are very unlikely under a statistical model of honest behaviours.

the first digits of the financial numbers before introducing errors. They then show that only after introducing non-zero mean errors to the dataset do they see deviations; and the larger the error introduced, the larger is the deviation from the law. The main consensus is that accounting-related data are expected to adhere to the Benford distribution and as deviation from the Benford distribution increases, the degree of errors increases. This property is akin to the idea of "hard-to-forge" signatures (Kossovosky, 2015 p.109). Only unique individuals can forge signatures, similarly only unique individuals could forge a dataset to follow Benford's Law, implying that any deviations from the theoretical frequency are likely to be due to errors or manipulation.

2. Data

We use the Third Sector Research data deposited in the UK Data Services by Alcock and Mohan (2015). The collection constitutes the largest dataset on organisational and financial characteristics of UK-based charities, co-operative organisations and mutual societies. Apart from standard items in the financial reports, the dataset provides detailed financial information on numerous types of expenses such as charitable and fundraising activities, voluntary incomes, administrative expenses, and employment statistics. Alcock and Mohan (2015) describe the collection process and provide descriptive analysis of the data.

The data are collected in five phases (TSRC 07-08; and 4 phases of Almanac 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).[15] We first convert all reported financial items to Sterling using exchange rates in the respective year.[16] We remove charities with negative total assets and negative spending on governance or charitable activities.[17] For simplicity and objectivity, we use all the financial information that appears in the balance sheet, income statements and cash flow statements to calculate our proxies. There are 135 financial items per annual report, however many are recorded as zeros. To ensure the comparability across non-profits and to improve the precision of the measures, we aggregate each NPO's annual report over the available years to have a pool of at least 100 non-zero financial observations (see Nigrini, 2012 for statistical rationale behind the threshold of 100). The practice is not uncommon in digital analysis.[18] One implicit assumption of aggregating the data is that the financial transactions across years

[15] Due to the surveys' structures, only non-profits with the total income of at least £25000 are collected.
[16] There are UK charities whose headquarters are in the UK but operate abroad and choose to report in the local currency including euros, Thai baht, Singaporean dollars, US dollars, … This practice does not alter the conformity/deviation of the dataset due to the scaling invariance property of the Benford's distribution (see Morrow, 2014 or Hill, 1995 for proof).
[17] Because our theoretical predictions hinge on the assumption that governance is relatively high, it is natural to test the theory on a subsample trimming NPOs with the lowest spending on governance. We do various trimming level of the dataset in Section VI and obtain generally consistent results. To avoid drawbacks of mistakenly excluding outliers, we report the full sample in our main analysis.
[18] Amiran et al. (2015) through simulated analysis and comparing with existing measures of reporting quality show that non-fabricated annual financial statements, whether in aggregate, by year or by organisation-year, are expected to generally conform to Benford's Law. Nigirini (2011, chapter 17) also uses multi-year financial statements to demonstrate the applicability of aggregate data in assessing errors and frauds by digital analysis. Henselmann et al. (2012) showed that Benford's law is applicable to the aggregated data of single-company annual reports given their reports are free of errors and manipulations.

are drawn from a set of random samples, allowing the aggregated data without errors and frauds to conform to Benford's Law (following Hill's theorem). It is reasonable since an NPO's financial transactions in different years could be driven by economy-wise shocks that are independent of shocks in previous years. Using US data in charitable giving from 1921-2007, List (2011) finds that the percentage changes in giving to most major charitable areas follow percentage changes in the S&P 500, suggesting that the distributions of the figures of charity revenues resemble that of S&P 500 indices, which are considered independent across years. Although we are not aware of any other research to support this assumption, we proceed and acknowledge this potential drawback of our methodology.

We next remove any NPOs whose total number of non-zero financial items is less than 100.[19] We discuss the sensitivity of this threshold in Section VI.2. One shortcoming of this practice is that we ignore a selection bias that some non-profits may choose to strategically submit fewer non-zero entries in each annual report. Zero transactions in an annual report could be due to two choices of the NPO: either they choose not to participate in some activities that could generate non-zero financial entries or they choose to withhold information by recording losses and some expenses as zero. The former is not a concern. The omitted variables underlying the decision of participation are independent of manipulative behaviour (for example, they choose not to work in Education services because they do not have the particular skills).

Withholding data is not something we can address with this dataset; however, we argue it is not critical 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 other non-zero financial items to be manipulated.[20] Our proxies based on Benford's Law are likely to pick up these deviations 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 proxies. The variable aims to account for both the diversity of the NPO's activities and, potentially, the level of the NPO's intention to disclose their financial details. Third, we estimate a sample selection model for missing observations using Heckit to distinguish two decisions: (1) whether to report more than 100 non-zero financial items (or to be included in our analysis); and (2) the optimal degree of inaccuracy once at least 100 non-zero items are reported. Despite being tentative due to the lack of more detailed data, we obtain consistent results for the main predictions even after controlling for the non-reporting issue.

In line with aggregating financial data we construct measures for other characteristics by taking averages of each NPO's over the reported period (for example: average of NPO A's total assets during the period 2008-2011). Although averaging over the period does not provide information on individual NPO's behaviour in an individual year, again it is not an uncommon approach. Michalski and Stolz (2013), also citing the lack of detailed data, aggregate quarterly macroeconomic data of several countries

---

Kossovsky (2015, p.90) provides a detailed review of previous literature to suggest a collection of monthly/yearly data or the market aggregate (combining multiple companies) is also expected to conform with the law.

[19] In the raw dataset, there is no coding for missing observations. A recorded zero item can be interpreted as genuinely missing (no information) or zero transaction.

[20] There is no recorded item for balancing errors in the raw dataset.

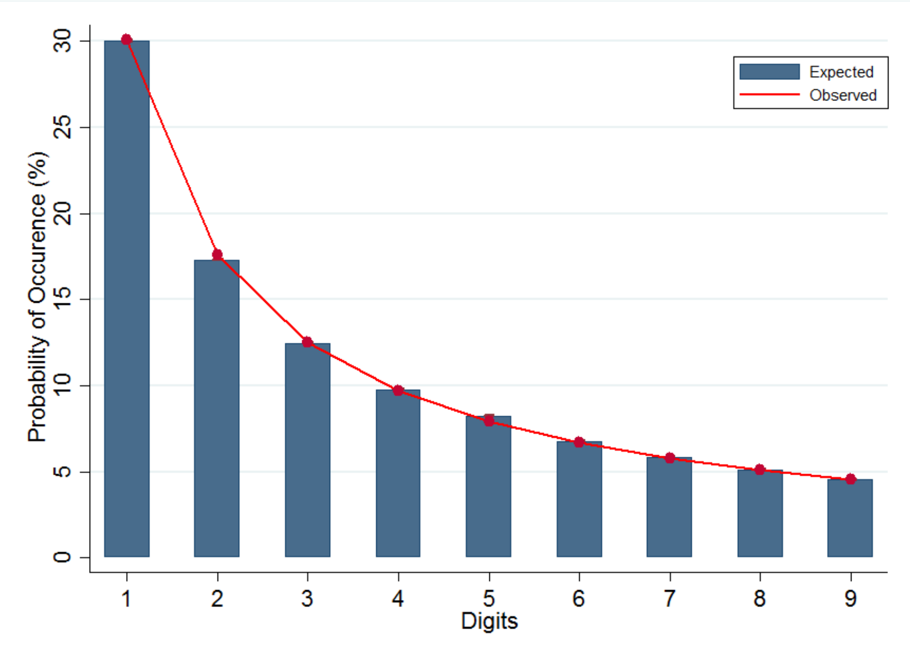
according to their economic characteristics. Using random subsampling to draw Bernoulli random subsamples from the aggregated data subsets, they show that their whole dataset adheres to Benford's Law, replicating the general expectation in Nye and Moul (2007). In another strand of economics, Matzkin (2013) provides a fuller discussion of the averaging method in non-parametric identification. We follow Matzkin's (2013) spirit as our focus is to explore the signs of the relationship derived from our proposed model, not the magnitudes of the results. We address this issue again in Section VI.2. The process yields a cross-sectional sample of 10,322 charities.

3. Proxies for reporting inaccuracy: deviations from Benford's Law

Regardless of the current consensus, the primary difficulty in applying Benford's Law to the detection of fraud is that many datasets do not naturally follow Benford's Law; and there is no definitive test to distinguish the two types. We assume that the accounting data of individual non-profits in the UK Third Sector, once free of errors and manipulations, would adhere to the Benford distribution. We graphically show that as a whole the UK Third Sector dataset conforms to Benford's Law. This result is in line with other accounting-related datasets in the corporate literature reviewed above and particularly that of Van Caneghem's (2015) study of the Belgian non-profit sector. As in previous studies, we emphasise that the conformity does not remove the possibility of individual NPO's financial reports containing frauds and errors.[21] We proceed and follow Amiran et al.'s (2015) simulated analysis to further assume that greater deviations from the Benford theoretical distribution reflect greater extents of errors and manipulation embedded in the data.

[21] Conformity acts as a stylised demonstration of the law's applicability in this source of data and the conformity in the whole sample may in fact be driven by the random errors or manipulations in independent magnitudes. Because different individual organisations have different and independent underlying mechanisms to generate errors or intention to manipulate; deviations from the expected distribution are likely to result from different sources with different magnitudes. Judge and Schechter (2009) also offer an intuitively similar argument whereby the process of aggregating the individual reports to a whole dataset can cause them to better fit the Benford distribution, even if the underlying data did not.

Figure 1. The UK Third Sector Research Data generally conform to the Benford distribution



Sources: Authors' calculation using the UK Third Sector Research Data by Alcock and Mohan (2015)

Measuring the extent that a dataset deviates from Benford's distribution has been debated in the digital analysis literature (see Morrow, 2014; Miller, 2015). Measures can be strongly influenced by the number of digits used, with some statistics requiring near-perfect conformity to the theoretical distribution as the number increases to not reject the null of conformity (Nigrini, 2012).[22] We employ the Mean Absolute Deviation (MAD) statistic in the main analysis. We complement this in Section VI.1 with three other "critical-value based" measures created from: (1) the Chi-square test statistics of goodness of fit, (2) the Kolmogorov – Smirnov (KS) statistics and (3) a binary variable 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). The MAD statistic is calculated as the mean of the absolute difference between the empirical proportion of each digit in each NPO's aggregated financial reports and their respective theoretical proportion according to Benford's Law (see Table 1):

$$\text{MAD} \equiv \frac{1}{9}\sum_{i=1}^{9} |P_o(d_i) - P_e(d_i)| \qquad (17)$$

where $d_i = 1, 2, \ldots, 9$ represents the first 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. Nigrinit (2012) shows that since the MAD statistic is independent of the pool of digits used, it becomes preferable to the other proxies when

[22] Barabesi et al. (2016) propose a testing procedure that claims to deliver exact significance levels and not rely on large-sample approximation.

examining larger pools of digits and comparing deviations of financial statements across organisations with different numbers of non-zero financial items reported. In addition, as there is no critical value involved in comparing the MAD across organisations, the statistic also provides a clear and objective measurement: the larger MAD statistic indicates further deviation from the theoretical distribution under the null hypothesis that the aggregated report is free of errors and manipulation.[23]

We also report the three critical-value based proxies for two reasons. First, we aim to demonstrate that our analysis is not sensitive to the choice of measures. Second, the three critical-value based proxies have been widely used by previous studies and practitioners (for example, Lin et al., 2014 and Michalski and Stolz, 2013 for the Chi-square test; Morrow, 2014, Amiran et al., 2015 for the KS test). They also offer ease of use and practical interpretations. Comparing the test statistics with a set of determined critical values can indicate whether to reject the null that the dataset conforms to the Benford distribution. Such practice, however, requires the users *ex ante* choice of critical values, removing the objectivity of the measurement. Nigrini (2011) discuss other deficiencies of using the critical-value based approaches, for example the dependence on assumptions of observational independence of the data or over-sensitivity to the number of digits used. Acknowledging these pitfalls, we derive the three complimentary proxies as follows:

$$D^2 \equiv N \sum_{i=1}^{9} \frac{[P_e(d_i) - P_o(d_i)]^2}{P_e(d_i)} \tag{18}$$

$$KS \equiv \max_{d_i \in \{1,2,\ldots,9\}} \left| \sum_{i=1}^{d_i} [P_o(d_i) - P_e(d_i)] \right| \tag{19}$$

$$Deviate = \begin{cases} 1 & \text{if } \Pr(KS \leq D_N(\alpha) = \frac{c(\alpha)\sqrt{2}}{\sqrt{N}}) < 0.05 \\ 0 & \text{if } \Pr(KS \leq D_N(\alpha) = \frac{c(\alpha)\sqrt{2}}{\sqrt{N}}) \geq 0.05 \end{cases} \tag{20}$$

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, $c(\alpha)$ is the Benford specific critical value at α calculated in Morrow (2014). Normally $\alpha = 0.05$ and $c(0.05) = 1.48$. Here, D^2, KS, $Deviate$ represent the Chi-square statistic, the KS statistic and the binary variable taking value one if we reject the null of that the underlying observed distribution of the NPO's aggregated financial report follows the Benford distribution at the 5% significance level and zero otherwise. Similar to the MAD statistic, greater values of the indices indicate that the tested data diverge further from the Benford distribution, hinting towards errors and manipulation.

Caveats of our proxies: These measures hinge on the assumption that an accurate financial statement dataset adheres to the Benford distribution; while manipulated and/or erroneous data deviate from the

[23] Nigrini (2011) recommends a table of "critical values for rejecting conformity" for practitioners. However, it is based on simulated datasets of specific dataset types.

law. Our task is also complicated by the fact that the NPO may not cheat all the time (serial cheaters might be found out quickly); so the proxies based on aggregate data cannot pinpoint for which year or which financial items the illicit behaviour has occurred. The proxies also cannot detect other types of cheating. First, by using the leading digits in constructing the measures, we focus on the most egregious type of manipulation, namely, that the reporters systematically alter the data. There are subtler ways in which organisations can misreport, such as rounding up numbers (which only affects the last digit), petty manipulation which also only affects the last digits of the data (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). Second, one NPO could change all its items by a common factor or in a creative way that preserve the Benford first-digit distribution. Although we offer some theoretical remedy by the incremental governance cost, we are unsure of the severity of this affecting our results. We, however, argue that it is not critical because measurement errors in the dependent variable do not lead to biased estimators. The only consequence is less precision in the estimated coefficients and lower t-statistics (Hausman, 2001). Third, the deviations from the Benford distribution could be due to poor data collection/book keeping (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 once we use the first-digit distribution. 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 book keeping and data maintenance, the first-digit distribution of NPOs with lowest spending on governance activities (such as accounting or auditing) should be affected the most. The data does not support this concern: for 35% of NPOs (1056) in the bottom 10% in terms of spending on governance activities, we fail to reject that their financial reports do not adhere to the Benford distribution according to the KS tests at 5% level of significance. Third, we test in Section VI.3 the sensitivity of our results to the level of spending on governance activities. Our results remain consistent when we exclude NPOs in the bottom 10%, 25%, and 50% in terms of spending on governance. We also conjecture that if poor book keeping resulted in fewer data points being collected, including these NPOs could alter our results. We find that our results hold when we alternate the threshold of the number of non-zero financial items used.

IV. Econometric methodology

1. Empirical specification

Denote b_i the optimal degree of inaccuracy chosen by NPO i. The key implications we test in the empirical analysis concern the effects of the choice of action (denoted a_i) and governance characteristics on a measurement of information manipulation (denoted g_i). We capture the effect of the thresholds and interconnection between a_i and g_i specified in Section II.3 through an interaction term in the following simple specification:

$$b_i = \gamma_0 + \gamma_1 a_i + \gamma_2 g_i + \gamma_3 \mathrm{I_i} + \gamma_4 X_i + \varepsilon_i \qquad (16)$$

where $\mathrm{I_i} = a_i \times g_i$ is the interaction between the optimal choice of action a^* and the exogenously set governance structure g_i; $\gamma_j, j \in \{1,2,3\}$ are the parameters of interest; X_i and γ_4 are respectively a set of other observable explanatory variables and its vector of parameters; γ_0 represent a constant and ε is an error component capturing unobservable NPO characteristics.

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

Main variables of interest: We use the MAD statistic to measure the amount of embedded manipulation/errors b_i as the dependent variable. We use the ratio of spending on charitable activities to the NPO's total income (*Charitable Spending)* to proxy for each NPO's choice of the optimal charitable action. Although the amount of income spent on charitable activities is not always a perfect signal to assess the charitable effort, it is highly correlated. A non-profit with a higher ratio could be inferred to be exerting higher effort in maximising the use of their income. We discuss the possibility of endogeneity in Section IV.2. For the weight of governance structure imposed on each NPO, g_i, we use the ratio of spending on governance activities (including administrative expenses, auditing/accounting fees and other relevant governance costs) to the total income (*Governance Structure)*. It aims to capture the external, observable governance structure of each NPO. The interaction term (*Interaction Term)*, I_i, is generated by simply multiplying *Charitable Spending* and *Governance Structure.*

Control variables: The set X_i aims to control for other observable organisation characteristics and potential determinants that affect the precision of our manipulation proxies. We include the log of total assets (*NPO size*) to control for size; and age by years NPO has been in operation (*Age)* to measure the NPO's establishment or familiarity with the sectoral norm (a standard practice, see Yetman and Yetman, 2011).[24] We also include the reported number of volunteers (*Volunteers*). Non-profit organisations are often overseen and run solely or largely by volunteers. In many cases, NPOs operate with modest internal accounting practises with volunteers serving as part-time bookkeepers (Keating and Frumkin, 2003). As the volunteers may receive little instruction or simply may not be as fully committed as they are expected to be, these deficiencies in training and dedication can limit the ability of non-profits to maintain an adequate control of their reporting accuracy. In contrast, having attracted a substantial base of volunteers could be a signal of the non-profit's strength of its philanthropy arms and concern about reputation. The consequence is the organisation becomes open and more transparent in their financial reports to maintain their position (see Corollary 3). The ultimate impact of volunteers depends on the balance between these

[24] The results are robust to using unlogged or squared amounts of total assets.

two arguments. Although excluding this variable does not alter the core results, the *Volunteers* variable is critical for our traditional IV approach.[25]

We include six binary variables (Yes = 1, No = 0) that capture whether the NPO: (1) reports expenditure on either internal or external audits (*Being audited)*, (2) receives grants from any local, national or foreign government (*Receive government grants)*, (3) reports zero fundraising expenses (*Zero fundraising*), (4) reports any losses from their investments/pension funds (*Losses from investments*), (5) receives restricted income (*Receive restricted income*), and (6) receives endowment funds (*Have endowment funds*). Previous non-profit studies indicate that the first three variables are expected to be associated with misreporting activities. Having the reports audited is a popular potential determinant of accounting fraud. We expect that being audited reduces the irregularity of the data. Dependence on some specific types of donations, particularly from governments, can lessen the non-profit's incentive to undertake illicit activities (see Garven et al., 2016). It also seems implausible that a non-profit could incur exactly zero expense in fundraising, hence reporting zero fundraising amounts suggests some reporting inaccuracy (see Krishnan et al. 2006 for the first to use this variable).

We include the last three control variables as potential determinants of misreporting. Incurring losses from investments or pension funds could induce the NPO's manager to manipulate their reports to hide the loss. We conjecture that receiving restricted income and endowment funds also affects the NPO's motivation for reporting manipulation. Restricted income is a well-established feature of charity finance and defined as a source of income from donations/gifts for a specific purpose but still within the charity's overall objectives. Provided that the trustees can still exercise discretion over the funds given, the charity upon receiving restricted income and endowment funds has ownership rights and will be acting as a principal instead of as an agent in the case of conduit giving. Despite limited literature on this conjecture, we expect that as the charity has more power over their restricted income and greater reputation concerns for future receipts, they have more incentive to behave diligently and report more accurate accounts. Another reason is that non-profits with endowments or more reliable funding sources are more likely to operate more sophisticated accounting systems that may be less prone to inaccuracy. To avoid any spurious relationship, we emphasise that the results remain generally similar if we replace these last two binary variables by the ratio of restricted income/ endowment funds to the total income (dependence on restricted income/ endowment funds) or exclude the variables from our specification.

Another pitfall of using the binary variables is that we lose information on the magnitudes of the respective financial items from the actual (reported) continuous values. Because these reported continuous values could have been manipulated already (for example underreporting losses or over-reporting gains or restricted income), using the binary variables is preferred to avoid possible measurement errors in multiple explanatory variables that would intractably bias our estimates (Hausman,

[25] In the absence of a control for labour provision and reputation concern of a charity, our proposed instruments based on the number of headcount staff and the social security cost may not satisfy the exclusion restrictions as the instruments may capture the effects of volunteers on the charity's accounting activity.

2001). As discussed in Section III.2, it remains the case that some NPOs could record zero losses even when they experienced losses from investments or pension funds. Since our binary variables are for the cumulated period (ever reporting losses), we conjecture that a more common behaviour would be to underreport losses rather than recording serially zero losses over the years. The consequence is that our binaries become less prone to measurement errors and can be assumed exogenous throughout. We also provide a partial remedy for the non-reporting issue below.

We include the number of non-zero financial entries (*Number of non-zeros*) and the number of yearly reports used (*Number of yearly reports*). The former is to control for the size of the pool of digits that potentially influences the statistical precision of the proxies. The latter is to control for the fact that NPOs who submit more annual reports can have more non-zero financial items even if they report fewer observations by year. As discussed in Section III.2, we include these proxies also to capture two other important aspects. First, more non-zero transactions could indicate a more diverse, complex structure of the NPO (for example, more charitable activities to expend resources), and such sophistication could affect the degree of manipulation or human error embedded in the reporting process. Second, we aim to partially account for the potential issue of non-reporting when some NPOs could strategically withhold information by recording the transactions as zeros. We conjecture that the more non-zero financial items reported, once we control for the NPO's size and the number of yearly reports, could be an indication for the NPO's openness towards transparency. For that reason, we expect *Number of non-zeros* to have a negative effect on the extent of manipulation and errors.[26] We assume that the whole set of control variables X_i are exogenous and proceed with our empirical strategy.

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

When estimating the parameters of Equation (16) by OLS, our main concerns are measurement errors and omitted variable bias that may influence both the manipulation extent (*MAD statistic*) and the optimal action choice (*Charitable Spending*). Although we attempt to minimise the risk of omitted variable bias in the error term ε_i by controlling for a set of potential confounders, there are two remaining issues. It is possible that an altruistic and able NPO, or its honest managers, would be more likely to spend more on charitable activities, and at the same time be less likely to engage in manipulative activities. The theory in Section II also suggests that there are confounding variables (the NPO's risk aversion and variance of organisation-specific risk) that correlate with both the optimal action choice and the degree of inaccuracy. A reverse causality is also possibly present. An NPO after receiving grants/donations may choose to misreport as their optimal decision and then divert the resource away from charitable spending to illicit activities (for example: creative accounting or for personal use). The observed level of charitable spending would be caused by the level of manipulation that the NPO chose to follow. Another potential

[26] Our main results remain generally consistent for all the four manipulation proxies even when we exclude these two control variables.

bias is measurement error in the regressors: because an NPO has an incentive to manipulate the reported amount of charitable spending, the value of the *Charitable spending* variable we observe could have been manipulated and therefore classified as mis-measured.[27] Coefficients and standard errors obtained from OLS are then biased and inconsistent. We deal with the endogeneity of *Charitable spending*, and potentially the *Interaction Term*, by two strategies: the first follows the traditional instrumental variable approach; while the second uses heteroscedasticity to estimate the parameter of interest as proposed by Lewbel (2012).

The traditional IV approach requires a set of instruments to isolate the effect of *Charitable Spending* on the manipulation amount proxied by the *MAD statistic.* Valid instruments need to satisfy two criteria: being strongly correlated with the endogenous variable(s) (weak identification) and orthogonal to the outcome variable after controlling for other potential confounders (exclusion restrictions). Because one potential source of endogeneity is measurement error caused by manipulation captured in the outcome variable, the valid instruments also need to be not subject to possible misreporting. Otherwise, the first stage estimates are biased and inconsistent, and the instruments themselves become correlated with the outcome. Based on the data availability, we propose two alternative instruments: the NPOs number of staff (*Headcount of staff*), and the actual spending on social security benefits (S*ocial security spending*). We argue that these two instruments are less likely to be misreported by NPOs. The number of official staff 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 less likely to falsify these figures. The instruments are also likely to satisfy the other two criteria. The number of staff and the amount of income spent on social security are expected to be positively correlated with the amount spent on charitable activities because more activities or services would require more paid employees, at least in the roles of supervision or programme planning. One concern that can invalidate our strategy is that the instruments may be correlated with the error term in explaining the quality of the NPO's report, hence not meeting the exclusion restriction. We argue that it is not the case: the manipulation indices here are calculated by statistical procedures and if the reports were manipulated, only a few specific staff such as the audit committee or the accounting division would have been involved. To our knowledge, there exists virtually no literature to indicate the influence of the size of employment on the accuracy and incidence of accounting errors. Popular ideal predictors that are relevant with employment aspects would have been executive salaries (Keating et al., 2008), size of committee board, presence of audit committee or the use of professional external accountants (Krishnan and Yetman, 2011; see Garven et al., 2016 for a fuller review). As we already control for the NPOs size, importance of administrative/governance, auditing and

[27] The incentive is strong as the ratio of charitable spending to the total income is one of the most commonly used metrics for evaluating efficiency and effectiveness in the non-profit sector (Krishnan et al., 2006) and research shows that higher charitable spending ratios are important to donors and associated with higher donations (see Garven et al., 2016 for a full review). Hofmann and McSwain (2013) review the current non-profit literature and acknowledge this endogeneity problem remains unaddressed.

volunteering, we expect that both the instruments can be excluded from our equation of interest in (16).[28] We further report Sargan-Hansen tests of overidentification to support our argument. As the *Interaction term* is possibly endogenous; we interact the proposed instruments with *Governance structure* to construct two additional instruments. The results remain similar if we control for the *Interaction term'* endogeneity, though the Hausman endogeneity test fails to reject its exogeneity.

If the justification for the two instruments being strong remains unconvincing, we adopt the limited information maximum likelihood (LIML) estimations to account for the possibility that weak instrumentation is present (see Murrey 2006 and Hansen et al., 2008). To check the robustness if the exclusion restriction does not hold, we use the second IV approach proposed by Lewbel (2012) that does not rely on the standard exclusion restriction.

Lewbel's (2012) two-step estimator exploits heteroscedasticity and higher moment conditions to construct internal instruments from the model's data without the need of any external source of variables. The idea of using heteroscedasticity in identification dates back to Wright's (1928) pioneering paper, and later developed by Rigobon (2003) and Lewbel (2012). Lewbel (2012) shows that two conditions are sufficient for identification without imposing the standard exclusion restriction. First, the error terms in the first stage regression is heteroscedastic. Given it is satisfied, the estimates are identified if there exists 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), this condition is normally satisfied in many models of endogeneity or mismeasurement, in which error correlations are due to some unobserved common factor. The reporting manipulation context represents a valid setting as the main driving force of endogeneity discussed above is either the NPO's unobserved characteristics or mismeasurement error. Although Lewbel (2012) suggests that the approach may be used when external sources of instruments or other identification methods are not available, using higher moment conditions is likely to provide less reliable estimates and it is not known how robust the results are to misspecification. Another undesirable feature of this method is that we fail to acknowledge any economic intuition underlying the generation of these instruments. To address these concerns, we supplement the set of internal instruments derived from Lewbel's (2012) approach with our proposed instruments to improve the efficiency of the heteroscedasticity-based IV estimator. Supplementing external instruments can also allow Sargan-Hansen-type tests of the orthogonality conditions or overidentifying restrictions to be performed, which would not be feasible in the case of exact identification by our proposed instruments. We interpret the results obtained from Lewbel's estimator as a robustness check in case there is doubt over our proposed instruments' validity. As the method is not well-known (see Emran and Hou, 2013 for recent applications), we briefly describe the estimator's intuition.

Assume that the model of interest is: $Y_1 = X'\beta_1 + \gamma_1 Y_2 + \varepsilon_1$ (18) and the endogeneity problem of Y_2 emerges from $Y_2 = X'\beta_2 + \varepsilon_2$ (19), where X is a set of exogenous regressors. The traditional IV

[28] Including either or both the instruments in the main equation of interest does not result in significant estimates.

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 of exogenous variables X_i by OLS and save the residuals ε_{1i}.
ii. Regress *Interaction term* on the set of exogenous variables X_i by OLS and save the residuals, ε_{2i}.
iii. Form instruments $Z_{ij} = (X_i - \overline{X_\iota})\varepsilon_{ji}$ 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 because 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. Our proposed traditional instruments may also be added to improve the efficiency and avoid overidentification.

This approach provides consistent estimates for our parameters of interest. The critical condition of heteroscedasticity in the first stage regressions can be tested using Breusch-Pagan test for heteroscedasticity after (i) and (ii). The greater the degree of heteroscedasticity in the error processes, the higher will be the correlation of the generated instruments with the included endogenous variables and stronger first-stage identification. To verify the second condition of the generated instruments being exogenous, we report Hayashi C test of orthogonal conditions and overidentification.

Descriptive statistics: Table 2 provides a summary of the variables used. From 10,322 non-profits from the UK third sector surveyed, only 25% provide financial data that conform to the Benford distribution according to KS test at 5% significance level. The average spending on charitable activities and governance activities, are 76% and 9% of their total incomes respectively. There are several extreme cases at 770% and 200% of the total incomes. In the sensitivity analysis, the results hold when these extreme values are excluded. The average NPO size in terms of total assets for the UK sector is on average £9.5 million, however, the distribution is heavily skewed to the right. On average, the UK non-profits have 21 professional staff and 10 volunteers. Again the extreme cases are respectively 3,192 and 17,500. The results also hold when these extreme values are excluded. Some 87% of NPOs are audited and 38% receive government spending. Surprisingly, 54% of the NPOs report that they never spend on fundraising, which as Krishman et al. (2006) suggest may signal some reporting inaccuracy. Some 17% of them have some losses from investment, while 16% have endowment funds.

Table 2. Summary statistics for the UK Third Sector sample

VARIABLES	Mean	s.d	Min	Max
MAD statistic	0.038	0.014	0.005	0.124
D^2 (Chi-square test statistic)	35.44	25.17	1.328	326.2
KS (KS test statistic)	0.129	0.062	0.014	0.524
$Deviate$ (1 = Deviate, 0 = Conform)	0.751	0.432	0	1
Explanatory variables				
Charitable spending	0.763	0.315	0	7.797
Governance structure	0.087	0.185	0	2.013
Size (Total Assets, £Millions)	9.551	97.27	£70	8,547
Age	20.87	14.80	0.564	50.89
Volunteers (total number of volunteers)	9.717	249.8	0	17,500
Being audited (1 = Yes, 0 = No)	0.873	0.334	0	1
Receive government grants (1 = Yes, 0 = No)	0.384	0.486	0	1
Zero fundraising (1 = Yes, 0 = No)	0.542	0.498	0	1
Losses from investments (1 = Yes, 0 = No)	0.173	0.378	0	1
Receive restricted income (1 = Yes, 0 = No)	0.481	0.500	0	1
Have endowment funds (1 = Yes, 0 = No)	0.160	0.367	0	1
Number of non-zeros	200.7	62.56	100	406
Number of yearly reports	5.626	1.285	2	8
Headcount (total number of staff)	21.22	101.8	0	3,192
Social security spending (£'000)	77.294	361.013	0	15169
Observations	10,322			

Notes: The sample includes NPOs with at least 100 non-zero financial items available, non-negative total assets and non-negative expenditure on charitable activities and governance. *Charitable spending, Governance structure* are calculated as the proportions of income spent on the corresponding items. All calculations are made after taking averages of the financial items over the reported yearly financial statements. *Sources*: The UK Third Sector Research Data Collection, deposited by Alcock and Mohan (2015)

Table 3 provides t-tests of differences between the subsamples of those who deviate from and conform to the Benford distribution according to KS test at 5% significance level. As expected, the manipulation indices of those who deviate are significantly larger than those who conform. Spending on charitable activities, age, probability of losing from investment and the number of submitted financial reports are also significantly larger for those who deviate than for those who conform. The rest of the table suggests that there are no systematic differences between the two subsamples, including spending on governance, size, the probability of reporting zero fundraising, having endowment funds, and the number of volunteers or professional staff. We further note that spending on social security and headcount of staff are not systematically different which we interpret as support for using as instruments.

Table 3. Descriptive statistics by conformity to the Benford distribution based on KS test

VARIABLES	Deviate	Conform	Difference	t-statistics	P-values
MAD statistic	0.042	0.027	0.015***	52.282	0.000
D^2 (Chi-square test statistics)	41.598	16.835	24.763***	47.747	0.000
KS (KS test statistics)	0.15	0.065	0.086***	75.347	0.000
Charitable spending	0.77	0.741	0.029***	4.000	0.000
Governance structure	0.088	0.084	0.004	0.842	0.400
Size (Total Assets)	8.894	11.537	-2.643	-1.193	0.233
Age	21.546	18.842	2.703***	8.046	0.000
Volunteers	7.271	17.108	-9.837*	-1.73	0.084
Being audited	0.871	0.877	-0.005	-0.702	0.483
Receive government grants	0.364	0.445	-0.081***	-7.366	0.000
Zero fundraising	0.543	0.54	0.002	0.214	0.831
Losses from investments	0.179	0.153	0.026***	3.07	0.002
Receive restricted income	0.474	0.504	-0.030***	-2.655	0.008
Have endowment funds	0.159	0.165	-0.006	-0.74	0.460
Number of non-zeros	202.592	195.183	7.409***	5.208	0.000
Number of yearly reports	5.696	5.414	0.282***	9.67	0.000
Headcount	21.437	20.558	0.879	0.379	0.704
Social security spending	77599	76372	-1226.174	-0.149	0.8814
Observations	7,555	2,567			

Notes: *** p <0.01, * p < 0.1. Conformity is based on Kolmogorov – Smirnov (KS) tests of the observed distribution following the expected distribution. According to the tests at 5% significance level, the subsample "Deviate" ("Conform") contains NPOs whose observed first-digit distribution deviates from (conforms to) the Benford distribution; or $P_{KS} \leq 0.95$ ($P_{KS} > 0.95$) The reported t-statistics and p-values are for two-sided Wald tests on differences between the two subsamples' means. *Sources*: The UK Third Sector Research Data Collection from Alcock and Mohan (2015).

V. Empirical results

Table 4 shows robust results from OLS with full control variables, 2SLS with our proposed instruments, and Lewbel's (2012) estimator. The robust standard errors are reported.[29] To verify the validity of our instruments (*Headcount* and *Headcount* × *Governance structure, Social security spending* and *Social security spending* × *Governance structure*), and to check how sensitive they are to the specification, we first alternatively include either pair of instruments and test for the exogeneity of the *Interaction term*. If we fail to reject the null of *Interaction term* being exogenous, we remove it from the set of endogenous variables in the sequential specification. Lewbel's (2012) estimator is complemented with our proposed

[29] We also estimate two base-line OLS regressions: (1) with the three main variables (no control variables) and (2) with control variables but excluding *Number of non-zeros* and *Number of yearly reports.* To save space, we omit the results, which are similar to the reported tables. We also use Variance Inflation Factor analysis (AIF) to see if multicollinearity drives our results. Small condition indices (substantially lower than 10) indicates our specifications are not subject to multicollinearity.

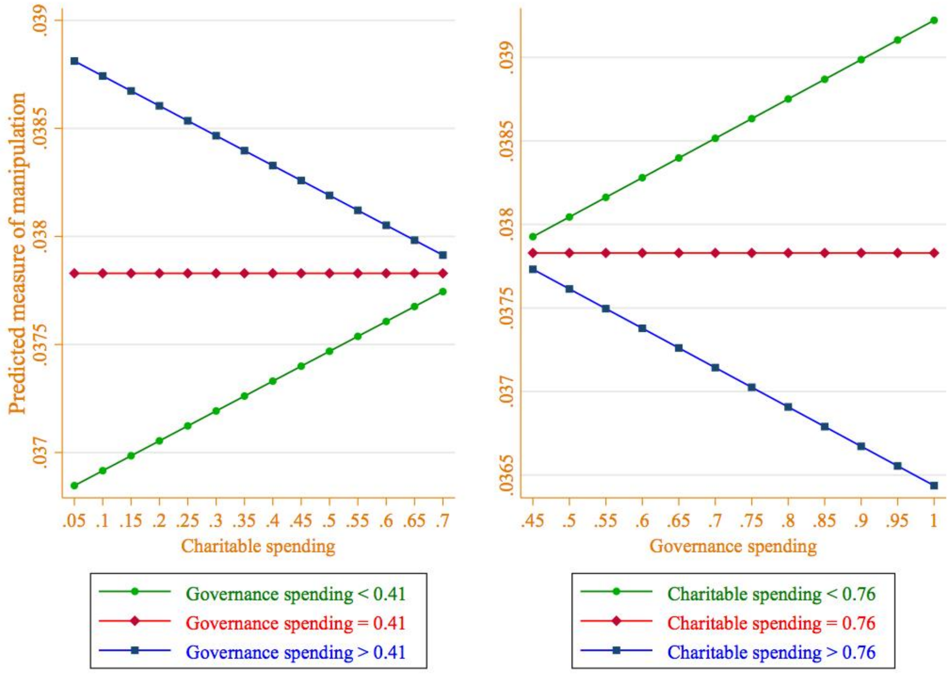
set of instruments. To explore the robustness of Lewbel's (2012) estimator to the inclusion of our instruments, we alternatively use different subsets of our instruments. We also use LIML estimators instead of 2SLS (or GMM) in all the estimators to rectify any potential weak identification. The unreported results are quantitatively unchanged. To interpret the effect of the thresholds, we calculate the point estimates of the thresholds for *Charitable spending* and *Government structure* as follows;

$$\tilde{a} = \frac{\text{Coefficient of } Interaction\ term}{\text{Coefficient of } Governance\ structure}$$

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

The estimates provide strong evidence to support our main theoretical predictions: the effect of charitable spending (a signal for the optimal amount of charitable effort) on the amount of reporting manipulation/errors depends on a threshold of the NPO's governance structure and vice versa, implying a trade-off between spending on charitable activities and governance. The evidence supports the signs predicted in Section IV.1. If governance spending exceeds a threshold (ranging from 15% to 40% of the total income, depending on the estimators), higher spending on charitable activities is correlated with a lower degree of irregularities detected by using Benford's Law. If the governance spending is lower than the threshold, higher charitable spending is correlated with a higher extent of information inaccuracy. The same interpretation is applied to spending on governance with the threshold of charitable spending ranging from 70%-75% of income. These findings are robust after accounting for various control variables and the endogeneity of *Charitable spending* and the *Interaction term*. We also experiment by treating *Government structure* as endogenous to see if the results alter. As there is no reliable instrument for the variable, we use Lewbel's (2012) estimator to undertake the estimation for brevity. After controlling for the potential endogeneity of *Government structure* by the heteroscedasticity-based IV approach, we still find similar estimates. As the respective Hausman test of endogeneity fails to reject the null hypothesis that *Government structure* can be treated as exogenous at 5%, we prefer to report our current empirical results. Figure 2 presents a graphical summary of the results using marginal effects estimated from OLS with full controls.

Figure 2. The marginal effects of governance spending and charitable spending on inaccuracy



Notes: Margin plot of the marginal effects obtained after OLS with full controls.
Sources: Authors' calculation using the UK Third Sector Research Data Collection.

Before further discussing the estimates, it is important to note that our proposed instruments and the Lewbel's (2012) estimator appear to work well in our data. Panel B in Table 4 shows that regardless of the instruments we reject the null hypothesis of the specification being under-identified (Kleibergen-Paap rk LM test), and instruments being weak (using Cragg-Donal Wald F statistics or Kleibergen-Paap rk F-statistics). Diagnostic tests for the first-stage estimations in Table 6 further support our proposed instruments and those generated by Lewbel's estimators are strongly correlated with the endogenous variables. Both the traditional F-statistics of excluded instruments and Sanderson-Windmeijer (2016) F-statistics are substantially large. We reject the null of homoscedastic errors in both first-stage equations using Breusch-Pagan/Cook-Weisberg test for heteroscedasticity at the $p < 0.01$ with high test statistics, suggesting strong correlations of the internally generated Lewbel's (2012) instruments. The Hansen's J statistics (Hayashi's C statistics for Lewbel's estimator) also support our arguments for exclusion restrictions. We fail to reject the null that the proposed instruments are orthogonal to the error terms in all cases. Table 6 presents Hausman tests for the endogeneity of *Charitable spending* and *Interaction term.* Although we can reject the null of exogeneity for *Charitable spending*, we fail to do so for *Interaction term.* This result indicates that it is not statistically necessary to control for the endogeneity of the *Interaction term*, suggesting that the specification 2SLS-6 (using *Headcount* and *Headcount* ×

Governance structure, Social security spending and *Social security spending* × *Governance structure* as instruments for *Charitable spending)* can be the preferred IV specification.

Discussion: We interpret the results as indicating that the current monitoring requirements could be counter-productive due to the trade-off between spending on charitable activities and spending to improve accountability. Our results provide an answer to the current call to put more pressure on the third sector governance structure. Theoretical and empirical results show that higher spending on governance may not guarantee better reliability of the non-profits' public financial records. If the non-profits spend less than a certain fraction of their income on charitable activities, spending more on governance may not improve the situation. The non-profits could take the higher pressure from donors, or the public, to place higher importance on their performance; or they could respond by inflating their reported value. The results are consistent with a cross-section study of the Ugandan NPO sector by Burger and Owens (2010). Using self-reported financial data, they measure NPOs' transparency by a binary variable of whether the NPO could provide the enumerator with their financial report that they claim is publicly available. Burger and Owens (2010) find that unrealistic requests by donors, measured by the frequency of updates and reports sent back to the funder, diminishes the level of transparency measured. We support their conclusion in the sense that over-spending on governance costs could be counter-productive in failing to motivate the organisations to adhere to a known reporting standard.

Spending larger fractions on charitable activities may also not signal more reliable financial reporting. The extent of reliability which we can infer from observing the NPO's recorded spending on charitable activities depends on the level of governance structure. Such structure needs to be above a certain level before we may infer that higher charitable spending is correlated with higher reliability of reports. It can be the case that the higher spending on charitable activities that we observe were inflated already. After we control for the potential endogeneity of the observed numbers, the conclusion remains unchanged.

Table 5 provides estimates for the control variables. Contradictory to some previous studies (such as Krishnan et al., 2006), size is positively correlated with the amount of irregularities. This may not be surprising. Keating and Erumkin (2003) suggest that as non-profits grow and new grants are received, the external demands for recording and processing information change. Because funding is often short-term, accounting systems then need modifying to meet current reporting needs, so that hybrid, manual accounting systems may be prone to errors. Our theory offers further explanation. Larger NPOs now face a trade-off decision: investing more to improve the accounting system to cope with the new load (normally need to spend more than some certain amount), or keep spending on activities to support their charitable aims. If they choose the former, relatively lower spending on charitable activities is correlated with lower reliability of their accounts, *ceteris paribus*. If they choose the latter, their error-prone accounting systems will deliver lower reporting accuracy.

Table 4. The amount of reporting inaccuracy and NPO's observable characteristics

VARIABLES	Dependent variable: MAD statistics							
	OLS with controls	2SLS-1	2SLS-2	2SLS-3	2SLS-4	2SLS-5	2SLS-6	Lewbel (2012)
Charitable spending	3.052***	9.720**	13.051***	14.169*	15.102**	11.708**	14.588***	2.439***
	[0.553]	[4.898]	[4.859]	[7.313]	[7.064]	[5.349]	[5.255]	[0.763]
Interaction term	-7.470***	-54.232**	-29.762***	-44.05***	-34.334**	-50.19***	-33.189***	-7.091***
	[2.748]	[21.945]	[11.398]	[13.522]	[16.254]	[16.159]	[12.258]	[2.587]
Governance structure	5.685***	28.539***	19.134***	25.78***	21.893**	27.53***	21.202***	5.161***
	[1.560]	[10.220]	[6.751]	[8.695]	[9.697]	[8.429]	[7.275]	[1.557]
Observations	10,322	10,322	10,322	10,322	10,322	10,322	10,322	10,322
R-squared	0.262	0.228	0.228	0.256	0.256	0.254	0.255	0.262

Panel B. Diagnostic Tests

	2SLS-1	2SLS-2	2SLS-3	2SLS-4	2SLS-5	2SLS-6	Lewbel (2012)
Underidentification Test: Kleibergen-Paap rk LM statistic chi2(1) (p-value)	71.76 (0.00)	77.14 (0.00)	8.961 (0.00)	106.9 (0.00)	104.2 (0.00)	115.5 (0.00)	351.7 (0.00)
Weak identification test (Cragg-Donald Wald F statistic)	23.36	31.97	20.99	27.53	14.41	18.06	249.4
Weak identification test (Kleibergen-Paap Wald rk F-statistic)	25.05	18.84	19.15	10.75	14.42	11.80	53.44
Stock-Yogo (2005) weak ID test critical values							
10% maximal IV relative bias	7.03	16.38	7.03	16.38	7.56	19.93	10.96
15% maximal IV relative bias	4.58	8.96	4.58	8.96	5.57	11.59	6.17
20% maximal IV relative bias	3.95	5.53	3.95	5.53	4.73	8.75	4.48
Overidentification test: Hansen J statistics (C test for Lewbel's 2012 estimator) (p-value)	0.00 (1.00)	2.55 [0.11]	0.00 (1.00)	1.16 (0.28)	1.00 (0.60)	3.85 (0.28)	6.27 (0.10)

Notes: p-values in parenthesis. Robust standard errors for arbitrary heteroscedasticity in bracket, *** p<0.01, ** p<0.05, * p<0.1
All estimates are scaled up by a factor of 1000 for the ease of interpretation.

2SLS-1: Use *Headcounts* and *Headcounts* × *Governance structure* as instruments
2SLS-2: Use *Headcounts* and *Headcounts* × *Governance structure*, treating *Interaction term* as exogenous

2SLS-3: Use *Social security spending* and *Social security spending*× *Governance structure* as instruments
2SLS-4: Use *Social security spending* and *Social security spending*× *Governance structure*, treating *Interaction term* as exogenous
2SLS-5: Use *Headcounts* and *Headcounts* × *Governance structure, Social security spending*× *Governance structure* as instruments.
2SLS-6: Use *Headcounts* and *Headcounts* × *Governance structure, Social security spending* and *Social security spending* × *Governance structure* as instruments, treating Interaction as exogenous.
Lewbel (2012): Using heteroscedasticity-based errors as instruments and using *Headcounts* and *Headcounts* × *Governance structure, Social security spending*× *Governance structure* to improve efficiency and avoid overidentification. All procedures are performed by GMM.

Underidentification test: Ho: matrix of reduced form coefficients has rank=K-1 (underidentified). Ha: matrix has rank=K (identified)
Weak identification test: Ho: equation is weakly identified. There is currently no critical values for Kleibergen-Paap Wald rk F statistic and Sanderson-Windmeijer (2016) statistics. Stock and Yogo (2005) have compiled critical values for the Cragg-Donald F statistic for the null of instruments being weak under i.i.d assumptions. Kleibergen-Paap Wald rk F statistic are correspondingly-robust statistics when i.i.d assumption is dropped. The critical values for this statistics are not yet available (Sanderson and Windmeijer, 2016)

Overidentification test for all instruments: Hansen J statistics: The joint null hypothesis is that the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.

Test of endogeneity: Ho: the specified endogenous can actually be treated as exogenous (Durbin-Wu-Hausman).
For Lewbel (2012), C statistic at chi2 (2) is reported: defined as the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments, where the suspect regressor(s) are treated as endogenous, and one for the equation with the larger set of instruments, where the suspect regressors are treated as exogenous. Unlike the Durbin-Wu-Hausman tests of endogeneity, this C-statistics test can report test statistics that are robust to various violations of conditional homoscedasticity (see Hayashi 2000, pp. 233-34).

Table 5. The amount of reporting inaccuracy and NPO's observable characteristics (*continued)*

CONTROL VARIABLES	Dependent variable: MAD statistics							
	OLS with controls	2SLS-1	2SLS-2	2SLS-3	2SLS-4	2SLS-5	2SLS-6	Lewbel (2012)
Size (logged total assets)	0.560***	0.730***	0.906***	0.922***	0.977***	0.815***	0.959***	0.556***
	[0.082]	[0.195]	[0.182]	[0.266]	[0.255]	[0.202]	[0.196]	[0.080]
Age	0.083***	0.078***	0.073***	0.072***	0.070***	0.075***	0.071***	0.084***
	[0.009]	[0.011]	[0.011]	[0.012]	[0.012]	[0.011]	[0.011]	[0.009]
Volunteers	-0.219	-0.230	-0.192	-0.199	-0.186	-0.217	-0.187	-0.161
(total number of volunteers)	[0.310]	[0.265]	[0.251]	[0.244]	[0.240]	[0.255]	[0.243]	[0.306]
Being audited (1 = Yes, 0 = No)	-2.481***	-2.368***	-2.512***	-2.466***	-2.519***	-2.410***	-2.517***	-2.463***
	[0.464]	[0.488]	[0.474]	[0.480]	[0.479]	[0.482]	[0.478]	[0.461]
Receive government grants	-1.994***	-1.965***	-1.991***	-1.981***	-1.991***	-1.972***	-1.991***	-1.981***
(1 = Yes, 0 = No)	[0.263]	[0.271]	[0.268]	[0.271]	[0.270]	[0.270]	[0.269]	[0.262]
Zero fundraising	-0.114	0.200	0.692	0.700	0.858	0.420	0.816	-0.122
(1 = Yes, 0 = No)	[0.270]	[0.533]	[0.479]	[0.675]	[0.633]	[0.536]	[0.506]	[0.271]
Losses from investments	0.370	0.432	0.433	0.448	0.446	0.440	0.443	0.396
(1 = Yes, 0 = No)	[0.346]	[0.353]	[0.357]	[0.360]	[0.363]	[0.355]	[0.360]	[0.345]
Receive restricted income	-0.615**	-0.785**	-0.717**	-0.766**	-0.738**	-0.778**	-0.733**	-0.570**
(1 = Yes, 0 = No)	[0.290]	[0.305]	[0.300]	[0.304]	[0.307]	[0.302]	[0.302]	[0.290]
Have endowment funds	-1.004***	-0.763*	-0.531	-0.505	-0.433	-0.649	-0.458	-1.116***
(1 = Yes, 0 = No)	[0.344]	[0.440]	[0.431]	[0.515]	[0.502]	[0.451]	[0.446]	[0.345]
Number of non-zeros	-0.172***	-0.181***	-0.191***	-0.191***	-0.194***	-0.186***	-0.193***	-0.172***
	[0.005]	[0.011]	[0.011]	[0.015]	[0.015]	[0.012]	[0.011]	[0.005]
Number of yearly reports	4.128***	4.462***	4.761***	4.799***	4.891***	4.611***	4.859***	4.085***
	[0.208]	[0.400]	[0.385]	[0.530]	[0.507]	[0.418]	[0.408]	[0.211]
Constant	40.701***	33.643***	28.318***	27.377***	25.779***	30.861***	26.414***	41.356***
	[1.415]	[6.385]	[6.112]	[9.227]	[8.820]	[6.804]	[6.615]	[1.496]

Notes: See above

Table 6. Summary of first-stage estimations

Diagnostic Tests	Estimators						
	2SLS-1	2SLS-2	2SLS-3	2SLS-4	2SLS-5	2SLS-6	Lewbel (2012)
First stage estimation for *Charitable spending*							
F test of excluded instruments [p-value]	25.496 [0.000]	18.837 [0.000]	70.121 [0.000]	79.292 [0.000]	60.554 [0.000]	55.949 [0.000]	51.100 [0.000]
(weak IV) Sanderson-Windmeijer (2016) F(1, 10309) [p-value]	43.860 [0.000]	18.837 [0.000]	157.736 [0.000]	79.292 [0.000]	74.678 [0.000]	55.949 [0.000]	58.310 [0.000]
(underidentification) Sanderson-Windmeijer Chi-square (1) [p-value]	43.924 [0.000]	37.732 [0.000]	157.966 [0.000]	158.83 [0.000]	224.403 [0.000]	224.186 [0.000]	1,521.964 [0.000]
(endogeneity) Wu-Hausman test F(1, 10308) [p-value]	6.608 [0.038]	3.623 [0.06]	11.15 [0.000]	3.390 [0.065]	10.402 [0.013]	4.158 [0.041]	4.323 [0.04]
Breusch-Pagan / Cook-Weisberg test for heteroscedasticity of H_0 errors are homoscedastic: Chi-square (1) [p-value]							2,075.01 [0.000]
First stage estimation for *Interaction term*							
F test of excluded instruments [p-value]	12.631 [0.000]	-	8.847 [0.000]	-	9.972 [0.000]	-	109.779 [0.000]
(weak IV) Sanderson-Windmeijer (2016) F(1, 10309) [p-value]	36.871 [0.000]	-	8.059 [0.000]	-	13.582 [0.000]	-	142.610 [0.000]
(underidentification) Sanderson-Windmeijer Chi-square (1) [p-value]	36.925 [0.000]	-	8.071 [0.000]	-	40.812 [0.000]	-	3,722.289 [0.000]
(endogeneity) Wu-Hausman test F(1, 10308) [p-value]	2.546 [0.111]	-	1.158 [0.282]	-	2.691 [0.101]	-	1.414 [0.234]
Breusch-Pagan / Cook-Weisberg test for heteroscedasticity of H_0 errors are homoscedastic: Chi-square (1) [p-value]							66,969.21 [0.000]

Notes: See Table 4 for explanations. Sanderson-Windmeijer (2016) F-statistics (Chi-square) are under the null of the instrument is actually weak (under-identified). It is used over the F-statistics when the i.i.d is dropped and there are at least two endogenous variables.

Table 7. Replications using other indices: the amount of reporting inaccuracy and NPO's observable characteristics.

VARIABLES	D^2 (Chi-square)			KS statistics			*Deviate* (1 = Deviate, 0 = Conform)		
	OLS	2SLS	Lewbel's	OLS	2SLS	Lewbel	OLS	2SLS	Lewbel's
Charitable spending	7.130***	20.360***	8.152***	16.355***	44.334***	20.860***	108.271***	225.851***	95.734***
	[1.171]	[3.143]	[1.581]	[2.536]	[8.290]	[3.501]	[18.022]	[65.450]	[23.685]
Interaction term	-11.785**	-45.028	-12.663***	-36.435***	-147.614**	-42.167***	-174.689**	-562.837	-102.206
	[5.038]	[27.684]	[4.247]	[11.818]	[74.527]	[12.065]	[71.961]	[503.416]	[74.311]
Governance structure	10.700***	30.127**	10.939***	30.592***	89.469***	34.486***	159.336***	372.345	138.010***
	[2.723]	[12.652]	[2.614]	[7.373]	[34.189]	[7.432]	[41.467]	[227.960]	[42.415]
Size (logged total assets)	1.630***	2.079***	1.701***	2.192***	3.065***	2.389***	18.059***	21.880***	18.348***
	[0.160]	[0.187]	[0.156]	[0.384]	[0.485]	[0.386]	[2.851]	[3.677]	[2.845]
Age	0.156***	0.143***	0.153***	0.310***	0.283***	0.303***	1.207***	1.089***	1.248***
	[0.018]	[0.019]	[0.018]	[0.043]	[0.044]	[0.043]	[0.306]	[0.314]	[0.303]
Volunteers	-0.000	-0.000	-0.001***	-0.002*	-0.002**	-0.002**	-0.028**	-0.028**	-0.035***
	[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]	[0.013]	[0.012]	[0.010]
Being audited	-2.264***	-2.289***	-1.994**	-9.458***	-9.340***	-9.263***	-28.470**	-28.303**	-27.914**
(1 = Yes, 0 = No)	[0.806]	[0.831]	[0.796]	[2.171]	[2.224]	[2.156]	[13.805]	[14.032]	[13.753]
Receive government grants	-3.065***	-3.058***	-3.145***	-8.035***	-7.985***	-8.029***	-40.973***	-40.831***	-40.842***
(1 = Yes, 0 = No)	[0.515]	[0.520]	[0.506]	[1.269]	[1.280]	[1.264]	[9.534]	[9.549]	[9.491]
Zero fundraising	0.285	1.325**	0.567	-0.809	1.105	-0.053	13.807	22.406**	13.431
(1 = Yes, 0 = No)	[0.523]	[0.591]	[0.526]	[1.284]	[1.481]	[1.293]	[9.361]	[11.081]	[9.435]
Losses from investments	-0.217	-0.131	-0.274	0.924	1.133	1.306	-0.647	0.177	3.016
(1 = Yes, 0 = No)	[0.759]	[0.766]	[0.745]	[1.657]	[1.671]	[1.649]	[12.614]	[12.644]	[12.510]
Receive restricted income	-0.788	-0.935	-0.615	-3.824***	-4.264***	-3.647***	-34.589***	-36.187***	-32.075***
(1 = Yes, 0 = No)	[0.578]	[0.590]	[0.569]	[1.403]	[1.432]	[1.399]	[10.423]	[10.556]	[10.355]
Have endowment funds	-2.926***	-2.308***	-2.948***	-5.592***	-4.381***	-5.561***	-31.709**	-26.437**	-32.996***
(1 = Yes, 0 = No)	[0.682]	[0.706]	[0.676]	[1.620]	[1.678]	[1.616]	[12.704]	[13.128]	[12.641]
Number of non-zeros	-0.146***	-0.170***	-0.153***	-0.486***	-0.533***	-0.504***	-0.718***	-0.924***	-0.725***
	[0.011]	[0.012]	[0.011]	[0.025]	[0.029]	[0.025]	[0.184]	[0.223]	[0.183]
Number of yearly reports	8.145***	8.972***	8.147***	12.899***	14.535***	13.367***	58.872***	65.965***	59.227***
	[0.422]	[0.468]	[0.423]	[0.981]	[1.117]	[0.992]	[6.883]	[7.980]	[6.849]
Constant	-8.761***	-25.005***	-9.639***	118.51***	85.690***	112.58***	257.796***	116.917	260.013***
	[2.800]	[4.482]	[3.007]	[6.645]	[11.918]	[7.214]	[46.965]	[91.868]	[49.474]

Notes: Robust standard errors for arbitrary heteroscedasticity in bracket, *** p<0.01, ** p<0.05, * p<0.1
All estimates except for D^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. The rest of results are generally similar and available upon request.

Having reports audited, receiving government grants, restricted income or endowment funds significantly reduce the amount of iregularities measured. These results are consistent with other studies and our previous conjectures. Reporting zero fundraising and losses from investments/pension funds, however, are not significantly correlated with worse reporting accuracy, despite hypothetical motivations for the NPO to falsify their data. The two control variables for the number of non-zero transactions being significant, once accounting for the number of reports, suggest that the number of non-zeros recorded affect our proxies with more non-zero financial items recorded being correlated with better accuracy. One reason could be that NPOs disclosing more non-zeros are those reporting truthfully. Although we cannot test this hypothesis with the current data, we offer a remedy in the next section.

VI. Sensitivity Analyses

We undertake four analyses to test the sensitivity of the results. We show that while the precise numerical magnitudes of the estimates of interest may vary depending on the subsample in use, the major conclusions that support our theoretical predictions are robust.

1. Sensitivity to the measurement of reporting inaccuracy

We first show that our results are not sensitive to the constructing algorithms of our measures. We re-estimate using the three critical-based alternatives: D^2 (Chi-square statistics), KS statistics, and the binary variable *Deviate* indicating whether the non-profit's data deviate from the Benford distribution. Table 7 reports the estimates from OLS with full controls, our preferred 2SLS (2SLS-6 in previous tables), and Lewbel's (2012) estimator.[30] Although we cannot directly compare the magnitudes of the coefficients, all the signs and significance are unchanged. For brevity, one can use the marginal effects obtained from estimates for *Deviate* to infer how observable characteristics impact the propensity to deviate from or conform to the Benford distribution.

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

One prevailing concern when using Benford's Law in digital analysis is to determine a sensible cut-off of the number of non-zero financial observations to include in the pool of data. We follow previous studies in using the threshold of 100 in the main analysis. Another lingering concern is that some non-profits could have withheld some information by recording zero financial transactions. By varying the cut-off, we explore how this concern affects our results. If the mechanisms underlying the decision to withholding information and manipulating the reported information are similar, we should not observe any systematic difference when we include more NPOS with more zero financial items, who are more likely to withhold information. We sequentially reduce the cut-off from 100 to 65 and re-do the analysis 35 times. The

[30] We also do an IV probit for the binary variable *Deviate.* As they are asymptotically similar, we report the simple 2SLS results for ease of interpretation (see Angrist and Pischke, 2009).

unreported results are quantitatively unchanged, suggesting that our results are not driven by the cut-off choice. Results are available on request.

3. Sensitivity to the sample in use

As the distributions of total assets and spending in the UK third sector are heavily skewed, one concern is that our results may be driven by extreme cases. 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), *Charitable spending,* and *Governance structure.* The main results are robust to trimming. Results are available on request.

4. Controlling for potential informational non-disclosure by Heckman model

Because we exclude NPOs with less than 100 non-zero financial items, there is a selection bias issue: the sample is restricted to those who report more than 100 non-zeros as we consider the excluded NPOs as missing observations. There are two reasons for having missing values. First, the excluded NPOs could have operated in simpler/fewer activities which generate no significant transactions. Second, they have strategically withheld information by recording some significant items as zero. Although we show above that our main results are not sensitive to the threshold of 100, we provide another piece of evidence using Heckman sample correction model. As the approach is well-known, we briefly describe the methodology as follows.

Let T be a binary variable taking value 1 if the NPO reports at least 100 non-zeros in our sample and 0 otherwise. The variable T also indicates whether the NPO is missing from our main analysis. We explore the selection of NPOs to record the observed number of non-zero financial items by running a probit regression of T on the set of explanatory variables $P(T = 1|X_i) = \Phi(\mathrm{X_i})$ where $\Phi(.)$ is the cumulative distribution function. Estimations from this regression can be used to predict the probability of each NPO reporting at least 100 non-zeros (and be included in the analysis). In the second stage, Heckit estimator corrects for self-selection by incorporating a transformation of the predicted individual probabilities in the first stage as additional explanatory variables (the inverse Mills ratios). Table 8 presents the results. Even when controlling for potential selection bias due to excluding NPOs who record less than 100 non-zeros, our results remain qualitatively unchanged for all the four indices. We also report the first-stage estimations to show determinants of the propensity to report at least 100 non-zeros. We report Wald tests of independence under the null that the two decisions can be taken independently. Although we reject the null for the MAD statistics, we fail to do so for the other critical-based measures. The exclusion restrictions are also satisfied for all the specifications: the number of volunteers appears to significantly increase the propensity of the NPO reporting at least 100 non-zeros throughout the period; while it has no significant impact on the main equation of interest. This result further supports our conjecture that it is the number of volunteers, not the number of professional staff, that have any influence on the accounting procedure of non-profits, either placing greater pressure to be transparent (but not necessarily accurate) or providing greater human resources for accounting tasks.[31] Another reason may be that NPOs with more volunteers are those who provide more services. We discount this explanation as we already account for the non-profit's size in the specifications. We further vary the threshold of 100 to 65 and obtain similar results.

Table 8. Heckit estimator for missing observations for the four indices

VARIABLES	First Stage	Second stage (degree of accuracy)			
		MAD	D^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 structure	0.26**	5.68***	10.92***	31.27***	163.69***
	[0.12]	[1.57]	[2.75]	[7.41]	[41.64]
Size (logged total assets)	0.27***	0.61***	1.60***	2.42***	18.20***
	[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 (1 = Yes, 0 = No)	8.40***	-2.37***	-2.81***	-9.29***	-34.24**
	[0.51]	[0.47]	[0.92]	[2.20]	[14.25]
Receive government grants (1 = Yes, 0 = No)	6.84***	-1.94***	-3.29***	-7.75***	-43.50***
	[0.54]	[0.26]	[0.55]	[1.27]	[9.67]
Zero fundraising (1 = Yes, 0 = No)	5.08***	-0.04	0.14	-0.65	13.11
	[0.48]	[0.27]	[0.56]	[1.29]	[9.46]
Losses from investments (1 = Yes, 0 = No)	4.91***	0.32	-0.25	0.71	-1.84
	[1.44]	[0.35]	[0.76]	[1.66]	[12.63]
Receive restricted income (1 = Yes, 0 = No)	9.86***	-0.55*	-0.95	-3.68***	-35.45***
	[0.85]	[0.29]	[0.60]	[1.41]	[10.49]
Have endowment funds (1 = Yes, 0 = No)	0.93	-1.03***	-2.97***	-5.76***	-32.24**
	[0.78]	[0.34]	[0.68]	[1.62]	[12.71]
Number of yearly reports	17.27***	4.25***	7.81***	13.32***	56.40***
	[0.53]	[0.21]	[0.49]	[0.99]	[7.39]
Number of non-zeros		-0.17***	-0.14***	-0.49***	-0.70***
		[0.01]	[0.01]	[0.03]	[0.19]
Observations	15,639	15,639	15,639	15,639	15,639
Chi-square (1)		6.66***	1.53	2.06	1.37
[p-value]		[0.01]	[0.22]	[0.15]	[0.24]

Notes: *** p<0.01, ** p<0.05, * p<0.1, robust standard errors are in brackets. First stage estimates probit of T (= 1 if included in the digital analysis by 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.

VII. Conclusion

This paper provides evidence on the trade-off between NPO spending on charitable activities and governance activities faced by both donors and NPOs using the UK Third Sector Data from 2007-2015. We first propose a theoretical model between a donor and an agent in a three-period reporting game to

[31] We replace/include the *Headcounts* variable into the first-stage equation and obtain insignificant estimates.

parameterise the optimal degree of irregularities embedded in self-reported financial reports. We construct four measures of reporting inaccuracy based on a mathematical phenomenon of the first-digit distribution, namely, Benford's Law. We find that financial figures from 25% of the sample do not conform to Benford's Law at the 10% significance level. We use alternative identification schemes to control for the potential endogeneity of the observed fraction of income spent on charitable activities, which is a measure of an NPO's optimal choice of charitable action. We rely on the exogeneity of the number of staff and the income spent on social security, which are publicly recorded and verified. Our second approach uses the heteroscedasticity-based estimator proposed by Lewbel (2012), which does not rely on the standard exclusion restrictions. We find that increased charitable spending (or programme ratio) leads to more accurate financial reports if the NPO spends at least 15%-40% of their income on governance activities. If the threshold is not met, the NPO may either exaggerate or neglect their reporting activities. On the other hand, increased accuracy of financial reports is associated with increased spending on governance activities only when the NPO already spends at least 70%-75% of their total income on charitable activities. If the threshold is not met, larger governance spending may put pressure on the organisation to misreport their data.

To conclude, we posit that nonstandard approaches to measure information irregularities, such as ours, have become inevitably necessary. Individuals who engaged in illicit behaviours actively attempt to avoid detections and uncovering their misdeeds usually involves using (potentially) misled information so that popular methods often fail to provide satisfactory results. Our study contributes to statistical analyses in a growing literature of forensic economics reviewed in Zitzewitz (2012). Our approach is a quick and easy tool for detecting potential irregularities, but we emphasise that it does not provide definite evidence for fraudulent behaviours, nor does it substitute auditing. Rather, we view the approach as a useful screening tool to identify potential organisations for further investigation. We advocate that the method could improve the efficiency in the allocation of the limited resources of regulatory bodies.

References

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

Anheier, H. K. (2009). What kind of nonprofit sector, what kind of society? Comparative policy reflections. *American Behavioral Scientist*, 52(7), 1082-1094.

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

Barabesi, L., Cerasa, A., Cerioli, A., & Perrotta, D. (2016). Goodness-of-fit testing for the Newcomb-Benford law with application to the detection of customs fraud. *Journal of Business & Economic Statistics*, (just-accepted).

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

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.

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.

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*. http://dx.doi.org/10.1111/1911-3846.12218.

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., 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.

Ebrahim, A. (2003). Accountability in practice: Mechanisms for NGOs. *World Development*, *31*(5), 813-829.

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.

Fassin, Y. (2009). Inconsistencies in activists' behaviours and the ethics of NGOs. *Journal of Business Ethics*, *90*, 503–521.

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.

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.

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.

Hansen, C., Hausman, J., & Newey, W. (2008). Estimation with Many Instrumental Variables. *Journal of Business & Economic Statistics*, *26*(4), 398-422.

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.

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.
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 organizations 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). Expense misreporting in non-profit organizations. *The Accounting Review*, *81*(2), 399-420.
Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, *30*(1), 67-80.
Lin, F., Wu, M., Fang, T. Y., & Wun, J. C. (2014). The relations among accounting conservatism, institutional investors and earnings manipulation. *Economic Modelling*, *37*, 164-174.
List, J. A. (2011). The market for charitable giving. *The Journal of Economic Perspectives*, 25(2), 157-180.
Matzkin, R. L. (2013). Nonparametric identification in structural economic models. *Annual Review Economics*, *5*(1), 457-486.
Miller, S. J. (Ed.). (2015). *Benford's Law: Theory and Applications*. Princeton University Press.
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.
Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *The journal of economic perspectives*, *20*(4), 111-132.
Morrow, L. (2014). *Benford's Law, families of distributions and a test basis. CEP Discussion Papers*, CEPDP1291. Centre for Economic Performance, LSE.
Nigrini, M. (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.
Rauch, B., Göttsche, M., Brähler, G., & Engel, S. (2011). Fact and Fiction in EU-Governmental Economic Data. *German Economic Review*, 12(3), 243-255.
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. *Advances in Economics and Econometrics. Ed. by Daron Acemoglu, Manuel Arellano, and Eddie Dekel*, *3*, 296-337.
Thakor, A. V. (2015). Strategic information disclosure when there is fundamental disagreement. *Journal of Financial Intermediation*, *24*(2), 131-153.
The Guardian, Singh, A. (2015). Transparency is great, but not at the cost of a charity's services. Retrieved September 05, 2016, from https://www.theguardian.com/voluntary-sector-network/2015/jan/08/transparency-great-cost-charity-services.
Steinberg, R. (2003). Economic theories of nonprofit organizations. In *The study of the nonprofit enterprise* (pp. 277-309). Springer US.
Van Caneghem, T. (2015). NPO Financial Statement Quality: An Empirical Analysis Based on Benford’s Law. *Voluntas*, 1-24.
Varian, H. (1972). Benford's Law (Letters to the Editor). *The American Statistician*, 26(3): 65.

Yetman, M. H., & Yetman, R. J. (2013). Do donors discount low-quality accounting information?. *The Accounting Review, 88*(3), 1041-1067.

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

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.

Wright, P. (1928). *The tariff on animal and vegetable oils*. New York: MacMillan.

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

Appendix A

Proof of Lemma 1

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

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

And

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

It implies that

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

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

Appendix B

Proof of Proposition 1

We first notice that

$$\left.\frac{\partial b^*}{\partial a^*}\right| \partial g = \frac{\partial b^*}{\partial g} \Big/ \frac{\partial a^*}{\partial g} \qquad (20)$$

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 $\frac{\partial a^*}{\partial g} > 0 \ \forall g \geq 0$ (21)

We can show that: $\quad \text{sign}\ \frac{\partial b^*}{\partial g} = \text{sign}\ T(g) \quad (22)$

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

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

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

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 2 shows that, given $g \geq 0$, sign $T(g) < 0$ if and only if $g \geq g_2$ and sign $T(g) > 0$ if and only if $g < g_2$.

Following (22), 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 (21), $\frac{\partial b^*}{\partial a^*}\Big| \partial g = \frac{\partial b^*}{\partial g} \Big/ \frac{\partial a^*}{\partial g} < 0$ if and only if $g \geq g_2$ and $\frac{\partial b^*}{\partial a^*}\Big| \partial g > 0$ if and only if $g < g_2$. Set g_2=$\tilde{g}$ the proof completes.

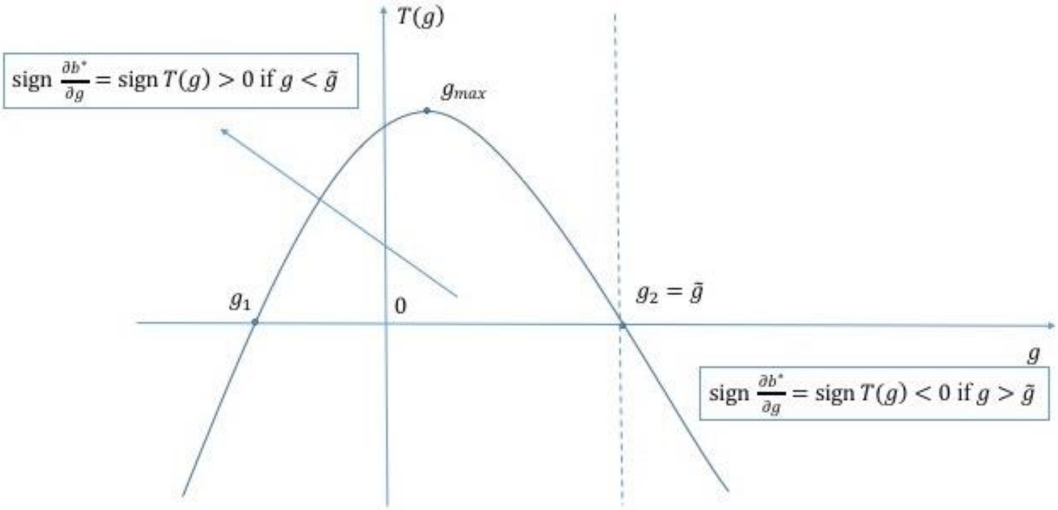


Figure 3. How sign $\frac{\partial b^*}{\partial g}$ and sign $T(g)$ behave when g varies in $[0, +\infty)$

Appendix C

Proof of Proposition 2

As $\frac{\partial a^*}{\partial g} > 0$ ∀g ≥ 0, $g \geq \tilde{g}$ if and only if $a \geq a_g = \tilde{a}$ and $g < \tilde{g}$ if and only if $a < a_g = \tilde{a}$ with $\tilde{g} = g_2$ specified as in Figure 2.

Figure 2 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.